ANNALS OF MATHEMATICS
Vol. 39, No. 3, July, 1938

CONTACT TRANSFORMATIONS

BY M. S. KNEBELMAN

(Received February 8, 1938)

1. Introduction

This note is concerned with the attempt to generalize a contact transformation as defined by Lie, in two natural directions; one may consider a geometrical transformation preserving contact of the m^{th} order between any two hypersurfaces—instead of first order;—and one may also consider transformations preserving contact of the first order between varieties of lower dimension than that of a hypersurface. What we shall prove is that both attempts fail; more specifically we shall show that a contact transformation of order m ($m > 1$) is *essentially* the $(m-1)^{\text{th}}$ extension of the ordinary contact transformation; and that a contact transformation between varieties other than hypersurfaces is *essentially* an extended point transformation.

A special case of the first theorem was proved by Vessiot in Actes Cong. Interbalkan Math., pp. 43–50 (1935).

2. Hypersurfaces with m^{th} order contact

Let V_n be the n-dimensional arithmetic space referred to a coördinate system x. Let S and S' be two hypersurfaces in V_n having the point x_0 in common. S and S' have contact of the m^{th} order at x_0 if—and only if—every curve through x_0 in one hypersurface has contact of the m^{th} order with the other hypersurface. It therefore remains to define m^{th} order contact between a hypersurface and a curve. Since we are concerned only with local properties of curves and hypersurfaces we assume that the point x_0 in question is not singular and that the functions involved are sufficiently regular in the neighborhood of x_0. Let a hypersurface S be given by $f(x) = f(x_0)$ and a curve C through x_0 by $x^i = x^i(t)$, where $\left(\frac{\partial f(x)}{\partial x^\alpha}\right)_{x_0}$ are not all zero and $x^i(t_0) = x_0^i$ and $\left(\frac{dx^i(t)}{dt}\right)_{t_0}$ are not all zero. We say S and C have contact of order m at x_0 if

$$\left(\frac{d^k f(x(t))}{dt^k}\right)_{t_0} = 0 \text{ for } k = 0, 1, \cdots, m, \quad \neq 0 \text{ for } k = m+1. \tag{2.1}$$

The individual equations in (2.1) depend on the equation chosen for the hypersurface and on the parametrization of the curve C. But the set of equations is

invariant under change of equation and change of parameter. Expanding (2.1) we get

$$\begin{aligned} &\frac{\partial f}{\partial x^\alpha}\frac{dx^\alpha}{dt} = 0, \\ &\frac{\partial^2 f}{\partial x^\alpha \partial x^\beta}\frac{dx^\alpha}{dt}\frac{dx^\beta}{dt} + \frac{\partial f}{\partial x^\alpha}\frac{d^2x^\alpha}{dt^2} = 0, \\ &\cdots, \end{aligned} \tag{2.2}$$

where the partial derivatives are evaluated at x_0 and the total derivatives at t_0.

If $g(x) = g(x_0)$ is the equation of a hypersurface S' having contact of m^{th} order with S, $\left(\frac{d^k g(x(t))}{dt^k}\right)_{t_0} = 0$, $k = 0, 1, \cdots, m$ for any curve C satisfying (2.1); that is,

$$\begin{aligned} &\frac{\partial g}{\partial x^\alpha}\frac{dx^\alpha}{dt} = 0, \\ &\frac{\partial^2 g}{\partial x^\alpha \partial x^\beta}\frac{dx^\alpha}{dt}\frac{dx^\beta}{dt} + \frac{\partial g}{\partial x^\alpha}\frac{d^2x^\alpha}{dt^2} = 0. \\ &\cdots, \end{aligned} \tag{2.3}$$

evaluated at x_0 and t_0, must be consequences of (2.2).

Let p_α, $p_{\alpha\beta}$, $\cdots$ $p_{\alpha_1\cdots\alpha_m}$ be a set of numbers of which we know that p_α are not all zero, the others being symmetric in all their subscripts; we call such a set of numbers the components of a contact element of order m. From (2.2) and (2.3) it follows that S and S' will have contact of m^{th} order at x_0 if

$$\begin{aligned} &g_\alpha = kf_\alpha, \qquad (k \neq 0), \\ &g_{\alpha\beta} = kf_{\alpha\beta} + k_\alpha f_\beta + k_\beta f_\alpha, \\ &g_{\alpha\beta\gamma} = kf_{\alpha\beta\gamma} + k_\alpha f_{\beta\gamma} + k_\beta f_{\gamma\alpha} + k_\gamma f_{\alpha\beta} + k_{\alpha\beta} f_\gamma + k_{\beta\gamma} f_\alpha + k_{\gamma\alpha} f_\beta, \\ &\cdots \\ &g_{\alpha_1\cdots\alpha_m} = kf_{\alpha_1\cdots\alpha_m} + \sum k_{\alpha_i} f_{\alpha_1\cdots\alpha_{i-1}\alpha_{i+1}\cdots\alpha_m} + \cdots. \end{aligned} \tag{2.4}$$

We thus see that p_α, $p_{\alpha\beta}$, $\cdots$, are *homogeneous* components of the contact element, homogeneity being defined by (2.4).

3. Contact transformation of order m

Let T_m be a non-singular point transformation in a space whose points are described by coördinates $(x^\alpha, p_\alpha, p_{\alpha\beta}, \cdots, p_{\alpha_1\cdots\alpha_m})$:

$$T_m \begin{cases} \bar{x}^i = \bar{x}^i(x, p_\alpha, p_{\alpha\beta}, \cdots p_{\alpha_1\cdots\alpha_m}) \\ \bar{p}_{i_1\cdots i_k} = \bar{p}_{i_1\cdots i_k}(x, p) \qquad\qquad k = 1, \cdots, m. \end{cases}$$

Let $f(x) = f(x_0)$ be a family of hypersurfaces in V_n such that in the neighborhood U_0 of x_0' there is one hypersurface of the family through each point x_0. This

family of hypersurfaces and T_m define a point transformation in U_0 which carries a point x_0 into $\bar{x}_0$ and the contact element $p(0)$ into $\bar{p}(0)$, given by

$$T_m^{(0)} \begin{cases} \bar{x}_0^i = x^i(x_0, p_0) \\ \bar{p}(0) = p(x_0, p_0). \end{cases}$$

Since x_0 is an arbitrary point in U_0,

$$\bar{x}^i = x^i(x, p(x)) \tag{3.1}$$

is a point transformation determined by T_m and $f(x)$. We assume that this transformation is non-singular, the condition on f being $|\Delta_\alpha \bar{x}^i| \neq 0$ where

$$\Delta_\alpha \varphi = \left(\frac{\partial}{\partial x^\alpha} + p_{\alpha\beta}\frac{\partial}{\partial p_\beta} + p_{\alpha\beta\gamma}\frac{\partial}{\partial p_{\beta\gamma}} + \cdots p_{\alpha\beta_1\cdots\beta_m}\frac{\partial}{\partial p_{\beta_1\cdots\beta_m}}\right)\varphi, \tag{3.2}$$

the p's being computed from $f(x)$. It should be observed that $\Delta_\alpha \bar{x}^i$ involves the contact element of order $m + 1$.

The point transformation (3.1) under the above restriction will carry the family of hypersurfaces $f(x) = f(x_0)$ into a family $\bar{f}(\bar{x}) = \bar{f}(\bar{x}_0)$ where

$$\bar{f}(\bar{x}) \equiv f(x)$$

in virtue of (3.1). However, the element $\bar{p}_{i_1\cdots i_k}(k = 1, \cdots, m)$ as defined by T_m will not in general be the contact element for $\bar{f}$. We say T_m is a contact transformation of order m if $\bar{p}$ is the contact element for $\bar{f}$ for *any* hypersurface subject of course to the restriction $|\Delta_\alpha \bar{x}^i| \neq 0$.

Since the variables p involved in T_m are not a unique set of components of the m^{th} order contact element, the functions involved in this transformation must be such that when x^α, p_α, $p_{\alpha\beta}\cdots$ are replaced by x^α, kp_α, $kp_{\alpha\beta} + k_\alpha p_\beta + k_\beta p_\alpha, \cdots$, $\bar{x}^i$, $\bar{p}_i$, $\bar{p}_{ij}\cdots$ are changed to an equivalent set of components. This means that $\bar{x}^i$ must be homogeneous of degree zero in the p's—homogeneity being defined by (2.4). We shall say that our transformation T_m is canonical if for any hypersurface f

$$\bar{p}_{i_1\cdots i_k} = \frac{\partial^k \bar{f}}{\partial \bar{x}^{i_1}\cdots\partial \bar{x}^{i_k}}, \qquad (k = 1, \cdots, m).$$

The transformation

$$\begin{aligned} \bar{x}^i &= x^i, \\ \bar{p}_i &= k(x, p)p_i, \\ \bar{p}_{ij} &= k(x, p)p_{ij} + k_i(x, p)p_j + k_j(x, p)p_i, \\ &\cdots, \end{aligned}$$

is a contact transformation which leaves every element invariant; hence geometrically it is an identity transformation and we shall show that any contact transformation T_m is the product of an identity transformation and a canonical

T_m. It is this canonical T_m that is the $(m-1)^{\text{th}}$ extension of the ordinary contact transformation.

The necessary and sufficient conditions for T_m to be a contact transformation may be obtained from the fact that

$$\frac{\partial \bar{f}}{\partial \bar{x}^i}\frac{d\bar{x}^i}{dt} = 0,$$
$$\frac{\partial^2 \bar{f}}{\partial \bar{x}^i \partial \bar{x}^j}\frac{d\bar{x}^i}{dt}\frac{d\bar{x}^j}{dt} + \frac{\partial \bar{f}}{\partial \bar{x}^i}\frac{d^2x^i}{dt^2} = 0,$$
$$\cdots,$$

must hold because

$$\frac{\partial f}{\partial x^\alpha}\frac{dx^\alpha}{dt} = 0,$$
$$\frac{\partial^2 f}{\partial x^\alpha \partial x^\beta}\frac{dx^\alpha}{dt}\frac{dx^\beta}{dt} + \frac{\partial f}{\partial x^\alpha}\frac{d^2x^\alpha}{dt^2} = 0,$$
$$\cdots,$$

where x and $\bar{x}$ are related by (3.1). Replacing $\frac{dx}{dt}, \frac{d^2x}{dt^2} \cdots$ by means of (3.1) we get

$$\bar{p}_i \Delta_\alpha \bar{x}^i = k p_\alpha,$$
$$\bar{p}_{ij} \Delta_\alpha \bar{x}^i \Delta_\rho \bar{x}^j + \bar{p}_i \Delta_\alpha \Delta_\beta \bar{x}^i = k p_{\alpha\beta} + k_\alpha p_\beta + k_\beta p_\alpha, \tag{3.3}$$

as the necessary and sufficient conditions on T_m. If we introduce the identity transformation

$$I_m \begin{cases} \bar{x}^i = \bar{x}^i & \bar{k}(\bar{x}, \bar{p}) = k(x, p) \\ \bar{p}_i = \dfrac{\bar{p}_i}{\bar{k}} & \bar{k}_i = \dfrac{k_\alpha \bar{\Delta}_i x^\alpha}{k} \\ \bar{p}_{ij} = \dfrac{\bar{p}_{ij} - \bar{k}_i \bar{p}_j - \bar{k}_j \bar{p}_i}{\bar{k}} & \end{cases} \tag{3.4}$$
$$\cdots,$$

we get a canonical transformation

$$C_m = I_m \cdot T_m,$$

where C_m may be singular, since I_m as a transformation of coördinates, may be singular. The conditions on C_m corresponding to (3.3) are

$$\bar{p}_i \Delta_\alpha \bar{x}^i = p_\alpha,$$
$$\bar{p}_{ij} \Delta_\alpha \bar{x}^i \Delta_\beta \bar{x}^j + \bar{p}_i \Delta_\alpha \Delta_\beta \bar{x}^i = p_{\alpha\beta}, \tag{3.5}$$

where $|\Delta_\alpha \bar{x}^i| \neq 0$ since $(\bar{x} = \bar{x})$ and we shall show that C_m is non-singular if T_m is. Operating on the first set of (3.5) with Δ_β we get

$$\Delta_\beta \bar{p}_i \Delta_\alpha \bar{x}^i + \bar{p}_i \Delta_\alpha \Delta_\beta \bar{x}^i = \Delta_\beta p_\alpha = p_{\alpha\beta},$$

so that the second set of equations are equivalent to

$$\bar{p}_{ij} \Delta_\alpha \bar{x}_j = \Delta_\alpha \bar{p}_i . \tag{3.6}$$

From (3.6) we get

$$\Delta_\beta \bar{p}_{ij} \Delta_\alpha \bar{x}^j + \bar{p}_{ij} \Delta_\alpha \Delta_\beta \bar{x}^j = \Delta_\beta \Delta_\alpha \bar{p}_i ,$$

treating x, p_α, $p_{\alpha\beta} \cdots$ as independent variables. Since $\bar{x}$, $\bar{p}$ are functions of x defined by the transformation (3.1) we get

$$\frac{\partial \bar{p}_{ij}}{\partial \bar{x}^k} \Delta_\beta \bar{x}^k \Delta_\alpha \bar{x}^j + \bar{p}_{ij} \Delta_\beta \Delta_\alpha \bar{x}^j = \Delta_\beta \Delta_\alpha \bar{p}_i ,$$

so that

$$\bar{p}_{ijk} \Delta_\alpha \bar{x}^k = \Delta_\alpha \bar{p}_{ij} , \tag{3.7}$$

and so in general we get

$$\begin{aligned} &\bar{p}_i \Delta_\alpha \bar{x}^i = p_\alpha , \\ &\bar{p}_{i_1 \cdots i_\lambda k} \Delta_\alpha \bar{x}^k = \Delta_\alpha \bar{p}_{i_1 \cdots i_\lambda} , \qquad \lambda = 1, \cdots, m-1, \end{aligned} \tag{3.8}$$

as the necessary and sufficient conditions in order that C_m should preserve contact of order m. Recalling that $\Delta_\alpha x^i$ involves $p_{\alpha_1 \cdots \alpha_{m+1}}$ equations (3.8) break up into two sets

$$\begin{aligned} &\bar{p}_i \frac{D\bar{x}^i}{Dx^\alpha} = p_\alpha , \qquad \bar{p}_i \frac{\partial \bar{x}^i}{\partial p_{\gamma_1 \cdots \gamma_m}} = 0 \\ &\bar{p}_{i_1 \cdots i_\lambda k} \frac{D\bar{x}^k}{Dx^\alpha} = \frac{D\bar{p}_{i_1 \cdots i_\lambda}}{Dx^\alpha} , \qquad \bar{p}_{i_1 \cdots i_\lambda k} \frac{\partial \bar{x}^k}{\partial p_{\gamma_1 \cdots \gamma_m}} = \frac{\partial \bar{p}_{i_1 \cdots i_\lambda}}{\partial p_{\gamma_1 \cdots \gamma_m}} , \end{aligned} \tag{3.9}$$

where

$$\frac{D}{Dx^\alpha} \varphi = \left(\frac{\partial}{\partial x^\alpha} + p_{\alpha\beta} \frac{\partial}{\partial p_\beta} + \cdots p_{\alpha\beta_1 \cdots \beta_{m-1}} \frac{\partial}{\partial p_{\beta_1 \cdots \beta_{m-1}}} \right) \varphi .$$

The commutator of the two operators $\dfrac{\partial}{\partial p_{\gamma_1 \cdots \gamma_m}}$ and $\dfrac{D}{Dx^\alpha}$ is given by

$$\left(\frac{\partial}{\partial p_{\gamma_1 \cdots \gamma_m}} \frac{D}{Dx^\alpha} - \frac{D}{Dx^\alpha} \frac{\partial}{\partial p_{\gamma_1 \cdots \gamma_m}} \right) \varphi = \delta^{(\gamma_1 \cdots \gamma_m)}_{(\alpha\beta_1 \cdots \beta_{m-1})} \frac{\partial \varphi}{\partial p_{\beta_1 \cdots \beta_{m-1}}} , \tag{3.10}$$

where $\delta^{(\gamma_1 \cdots \gamma_m)}_{(\beta_1 \cdots \beta_m)}$ is 1 or 0 according to whether $\gamma_1 \cdots \gamma_m$ is or is not a permutation of $\beta_1 \cdots \beta_m$. Applying this commutator to $\bar{x}^i$ we get, because of the first set of (3.9),

$$\delta^{(\gamma_1 \cdots \gamma_m)}_{(\alpha\beta_1 \cdots \beta_{m-1})} \bar{p}_i \frac{\partial \bar{x}^i}{\partial p_{\beta_1 \cdots \beta_{m-1}}} + \frac{\partial \bar{p}_i}{\partial p_{\gamma_1 \cdots \gamma_m}} \frac{D\bar{x}^i}{Dx^\alpha} - \frac{D\bar{p}_i}{Dx^\alpha} \frac{\partial \bar{x}^i}{\partial p_{\gamma_1 \cdots \gamma_m}} = 0,$$

and from (3.9) for $\lambda = 1$ we have

$$\frac{\partial \bar{p}_i}{\partial p_{\gamma_1 \cdots \gamma_m}} \frac{D\bar{x}^i}{Dx^\alpha} = \bar{p}_{ij} \frac{D\bar{x}^i}{Dx^\alpha} \frac{\partial \bar{x}^j}{\partial p_{\gamma_1 \cdots \gamma_m}} = \frac{D\bar{p}_i}{Dx^\alpha} \frac{\partial \bar{x}^i}{\partial p_{\gamma_1 \cdots \gamma_m}}.$$

Hence, since $\delta^{(\gamma_1 \cdots \gamma_m)}_{(\alpha\beta_1 \cdots \beta_{m-1})}$ is a non-singular matrix, we have

$$\bar{p}_i \frac{\partial \bar{x}^i}{\partial p_{\beta_1 \cdots \beta_{m-1}}} = 0. \tag{3.11}$$

From (3.9) for $\lambda = 1$ we get

$$\delta^{(\gamma_1 \cdots \gamma_m)}_{(\alpha\beta_1 \cdots \beta_{m-1})} \left(\bar{p}_{ij} \frac{\partial \bar{x}^j}{\partial p_{\beta_1 \cdots \beta_{m-1}}} - \frac{\partial \bar{p}_i}{\partial p_{\beta_1 \cdots \beta_{m-1}}} \right) + \frac{\partial \bar{p}_{ij}}{\partial p_{\gamma_1 \cdots \gamma_m}} \frac{D\bar{x}^j}{Dx^\alpha} - \frac{D\bar{p}_{ij}}{Dx^\alpha} \frac{\partial \bar{x}^j}{\partial p_{\gamma_1 \cdots \gamma_m}} = 0,$$

and for $\lambda = 2$ we have

$$\frac{\partial \bar{p}_{ij}}{\partial p_{\gamma_1 \cdots \gamma_m}} \frac{D\bar{x}^j}{Dx^\alpha} = \bar{p}_{ijk} \frac{\partial \bar{x}^k}{\partial p_{\gamma_1 \cdots \gamma_m}} \frac{D\bar{x}^j}{Dx^\alpha} = \frac{D\bar{p}_{ik}}{Dx^\alpha} \frac{\partial \bar{x}^k}{\partial p_{\gamma_1 \cdots \gamma_m}}.$$

Hence

$$\bar{p}_{ij} \frac{\partial \bar{x}^j}{\partial p_{\beta_1 \cdots \beta_{m-1}}} = \frac{\partial \bar{p}_i}{\partial p_{\beta_1 \cdots \beta_{m-1}}}. \tag{3.12}$$

By repeated application of this commutator process to equations (3.9) we reduce them to

$$\bar{p}_i \frac{\partial \bar{x}^i}{\partial x^\alpha} = p_\alpha, \qquad \bar{p}_i \frac{\partial \bar{x}^i}{\partial p_\alpha} = 0, \qquad \bar{p}_i \frac{\partial \bar{x}^i}{\partial p_{\gamma_1 \cdots \gamma_k}} = 0, \qquad k = 1, \cdots, m, \tag{3.13}$$

$$\bar{p}_{ij} \frac{{}_2D\bar{x}^i}{Dx^\alpha} = \frac{{}_2D\bar{p}_i}{Dx^\alpha}, \qquad \bar{p}_{ij} \frac{\partial \bar{x}^i}{\partial p_{\gamma_1 \cdots \gamma_k}} = \frac{\partial \bar{p}_i}{\partial p_{\gamma_1 \cdots \gamma_k}}, \qquad k = 2, \cdots, m, \tag{3.14}$$

$$\bar{p}_{ijh} \frac{{}_3D\bar{x}^h}{Dx^\alpha} = \frac{{}_3D\bar{p}_{ij}}{Dx^\alpha}, \qquad \bar{p}_{ijh} \frac{\partial \bar{x}^h}{\partial p_{\gamma_1 \cdots \gamma_k}} = \frac{\partial \bar{p}_{ij}}{\partial p_{\gamma_1 \cdots \gamma_k}}, \qquad k = 3, \cdots, m, \tag{3.15}$$

and so on, where

$$\frac{{}_\lambda D\varphi}{Dx^\alpha} = \frac{\partial \varphi}{\partial x^\alpha} + p_{\alpha\beta} \frac{\partial \varphi}{\partial p_\beta} + \cdots p_{\alpha\beta_1 \cdots \beta_{\lambda-1}} \frac{\partial \varphi}{\partial p_{\beta_1 \cdots \beta_{\lambda-1}}}.$$

The integrability conditions of (3.13) can be obtained in the form

$$(x^\alpha, x^\beta) = 0, \qquad (p_\alpha, p_\beta) = 0, \qquad (p_\alpha, x^\beta) = \delta^\alpha_\beta, \tag{3.16}$$

(u, v) being the Poisson parenthesis, and

$$\frac{\partial \bar{p}_i}{\partial p_{\gamma_1 \cdots \gamma_k}} \frac{\partial \bar{x}^i}{\partial p_{\alpha_1 \cdots \alpha_h}} - \frac{\partial \bar{x}^i}{\partial p_{\gamma_1 \cdots \gamma_k}} \frac{\partial \bar{p}_i}{\partial p_{\alpha_1 \cdots \alpha_h}} = 0, \quad k > 1, \quad h = 1, \cdots, m. \tag{3.17}$$

Let x^A stand for the variables $x^1, \cdots, x^n, p_1, \cdots, p_n$ and similarly for $\bar{x}^A$ and consider a linear transformation defined by $\left\| \frac{\partial \bar{x}^A}{\partial x^B} \right\|$, satisfying (3.16). Let ϵ_{AB}

be a tensor whose components in the $\bar{x}$ coördinate system are $\bar{\epsilon}_{AB}$, defined by the matrix $\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$ where I is the n rowed identity matrix. Then

$$\epsilon_{CD} = \frac{\partial \bar{x}^A}{\partial x^C} \frac{\partial \bar{x}^B}{\partial x^D} \bar{\epsilon}_{AB}$$

and because of (3.16) we find that ϵ is the same matrix. Hence $\left|\frac{\partial \bar{x}^A}{\partial x^B}\right|^2 = 1$ and the matrix $\left\|\frac{\partial \bar{x}^A}{\partial x^C}\right\|$ is non-singular. It therefore follows from (3.17) for $h = 1$ that

$$\frac{\partial \bar{x}^i}{\partial p_{\gamma_1 \cdots \gamma_k}} = 0 \qquad \frac{\partial \bar{p}_i}{\partial p_{\gamma_1 \cdots \gamma_k}} = 0 \quad \text{for } k > 1.$$

Therefore C_m is the $(m-1)^{\text{th}}$ extension of a homogeneous contact transformation and is non-singular for those values of $p_{\alpha\beta}$ for which

$$\left|\frac{\partial \bar{x}^i}{\partial x^\alpha} + p_{\alpha\beta} \frac{\partial \bar{x}^i}{\partial p_\beta}\right| \neq 0.$$

We therefore have

THEOREM 1. *The most general contact transformation of order m for hypersurfaces is the product of an identity transformation and the* $(m-1)^{\text{th}}$ *extension of a contact transformation of order* 1.

4. Contact of order one between varieties of dimension $n - r$

Let $f^\sigma(x) = f^\sigma(x_0)$, $\sigma = 1, \cdots, r$ be a set of equations defining an S_{n-r} through the point x_0, $\left\|\frac{\partial f^\sigma}{\partial x^\alpha}\right\|_{x_0}$ being of rank r. A curve C $x^i = x^i(t)$ through x_0 will have contact with S_{n-r} if $\left(\frac{df^\sigma(x(t))}{dt}\right)_{t_0} = 0$, that is, if $\frac{\partial f^\sigma}{\partial x^\alpha}\frac{dx^\alpha}{dt} = 0$. Since $F^\tau(f^\sigma) = F^\tau(f_0)$ with $\left|\frac{\partial F^\tau}{\partial f^\sigma}\right| \neq 0$ represents the same variety S_{n-r}, the contact element (of the first order) $p_\alpha^{(\sigma)}$ is determined to within a non-singular matrix k_τ^σ i.e. $p_\alpha^{(\sigma)}$ and $k_\tau^\sigma p_\alpha^{(\tau)}$ define the same geometrical element. We again consider a non-singular transformation

$$T^{(r)} \begin{cases} \bar{x}^i = \bar{x}^i(x, p^{(1)}, \cdots, p^{(r)}) \\ \bar{p}_i^{(\sigma)} = \bar{p}_i^{(\sigma)}(x, p^{(1)}, \cdots, p^{(r)}). \end{cases}$$

As in §3 we apply the transformation $T^{(r)}$ to a family of varieties $f^\sigma(x) = f^\sigma(x_0)$ such that

$$\bar{x}^i = \bar{x}^i\left(x, \frac{\partial f^{(1)}}{\partial x} \cdots, \frac{\partial f^{(r)}}{\partial x}\right) \tag{4.1}$$

is a non-singular point transformation, i.e. such that

$$\frac{D\bar{x}^i}{Dx^\alpha} = \frac{\partial \bar{x}^i}{\partial x^\alpha} + \frac{\partial^2 f^{(\sigma)}}{\partial x^\alpha \partial x^\beta} \frac{\partial \bar{x}^i}{\partial p_\beta^{(\sigma)}}$$

is a non-singular matrix. This point transformation will carry every such variety S_{n-r} into $\bar{S}_{n-r}$ given by $\bar{f}^\sigma(\bar{x}) = \bar{f}^\sigma(\bar{x}_0)$ where $\bar{f}^\sigma(\bar{x}) \equiv f^\sigma(x)$ because of (4.1). A curve C will be carried into $\bar{C}$ and $T^{(r)}$ will be a contact transformation if $\bar{C}$ and $\bar{S}$ have contact whenever C and S do. This means that $\bar{p}_i^{(\sigma)} \frac{d\bar{x}^i}{dt} = 0$ whenever $p_\alpha^{(\sigma)} \frac{dx^\alpha}{dt} = 0$ and since $\frac{d\bar{x}^i}{dt} = \frac{Dx^i}{Dx^\alpha} \frac{dx^\alpha}{dt}$ we must have $\bar{p}_i^{(\sigma)} \frac{Dx^i}{Dx^\alpha} = k_\tau^\sigma p_\alpha^{(\tau)}$ for arbitrary values of $p_{\alpha\beta}^{(\sigma)}$. Hence the necessary and sufficient conditions that T^r must satisfy to be a contact transformation are

$$\bar{p}_i^{(\sigma)} \frac{\partial \bar{x}^i}{\partial x^\alpha} = k_\tau^\sigma p_\alpha^{(\tau)}, \qquad \bar{p}_i^{(\sigma)} \frac{\partial \bar{x}^i}{\partial p_\alpha^{(\tau)}} = 0, \qquad |k_\tau^\sigma| \neq 0. \tag{4.2}$$

The integrability conditions of (4.2) are

$$\begin{aligned}
&\frac{\partial \bar{p}_i^\lambda}{\partial p_\beta^\nu} \frac{\partial \bar{x}^i}{\partial p_\alpha^\mu} - \frac{\partial \bar{p}_i^\lambda}{\partial p_\alpha^\mu} \frac{\partial \bar{x}^i}{\partial p_\beta^\nu} = 0, \\
&\frac{\partial \bar{p}_i^\lambda}{\partial x^\beta} \frac{\partial \bar{x}^i}{\partial x^\alpha} - \frac{\partial \bar{p}_i^\lambda}{\partial x^\alpha} \frac{\partial \bar{x}^i}{\partial x^\beta} = p_\alpha^\mu \frac{\partial k_\mu^\lambda}{\partial x^\beta} - p_\beta^\mu \frac{\partial k_\mu^\lambda}{\partial x^\alpha}, \\
&\frac{\partial \bar{p}_i^\lambda}{\partial p_\beta^\mu} \frac{\partial \bar{x}^i}{\partial x^\alpha} - \frac{\partial \bar{p}_i^\lambda}{\partial x^\alpha} \frac{\partial \bar{x}^i}{\partial p_\beta^\mu} = \delta_\alpha^\beta k_\mu^\lambda + p_\alpha^\nu \frac{\partial k_\nu^\lambda}{\partial p_\beta^\mu}.
\end{aligned} \tag{4.3}$$

Since the transformation T^r is non-singular, the equations

$$\begin{aligned}
d\bar{x}^i &= \frac{\partial \bar{x}^i}{\partial x^\alpha} dx^\alpha + \frac{\partial \bar{x}^i}{\partial p_\alpha^\lambda} dp_\alpha^\lambda, \\
d\bar{p}_i^\mu &= \frac{\partial \bar{p}_i^\mu}{\partial x^\alpha} dx^\alpha + \frac{\partial \bar{p}_i^\mu}{\partial p_\alpha^\lambda} dp_\alpha^\lambda,
\end{aligned}$$

must be uniquely solvable for dx^α and dp_α^λ. Multiplying the first of the above equations by $\frac{\partial \bar{p}_i^\mu}{\partial p_\beta^\nu}$ and the second by $\frac{\partial \bar{x}^i}{\partial p_\beta^\nu}$ and subtracting, we get because of (4.3)

$$\frac{\partial \bar{p}_i^\mu}{\partial p_\beta^\nu} d\bar{x}^i - \frac{\partial \bar{x}^i}{\partial p_\beta^\nu} d\bar{p}_i^\mu = k_\nu^\mu \left(\frac{\partial x^\beta}{\partial \bar{x}^i} d\bar{x}^i + \frac{\partial x^\beta}{\partial \bar{p}_i^\sigma} d\bar{p}_i^\sigma \right) + \bar{p}_i^\tau d\bar{x}^i \bar{k}_\tau^\rho \frac{\partial k_\rho^\mu}{\partial p_\beta^\nu},$$

when $\bar{k}$ is the matrix inverse to k. Since this must hold for arbitrary values of $d\bar{x}^i$ and $d\bar{p}_i^\mu$ we get

$$-\delta_\mu^\lambda \frac{\partial \bar{x}^i}{\partial p_\beta^\nu} = k_\nu^\lambda \frac{\partial x^\beta}{\partial \bar{p}_i^\mu}. \tag{4.4}$$

From (4.4) for $\lambda \neq \mu$ we have $k_r^\lambda \dfrac{\partial x^\beta}{\partial \bar{p}_i^\mu} = 0$ and since for a given λ, $k_1^\lambda, k_2^\lambda, \cdots k_r^\lambda$ can not all be zero it follows that $\dfrac{\partial x^\beta}{\partial \bar{p}_i^\mu} = 0$ and hence from (4.4) that

$$\frac{\partial \bar{x}^i}{\partial p_\beta^r} = 0. \tag{4.5}$$

Hence $\bar{x}^i$ are functions of x alone, and $\left|\dfrac{\partial \bar{x}^i}{\partial x^\alpha}\right| \neq 0$ so that the equations of T^r are

$$\bar{x}^i = \bar{x}^i(x) \qquad \bar{p}_i^\lambda = k_\mu^\lambda p_\alpha^\mu \frac{\partial x^\alpha}{\partial \bar{x}^i}.$$

We therefore have

THEOREM 2. *The most general contact transformation in V_n for varieties of dimension $n - r$ $(r > 1)$ is the product of an identity transformation and the first extension of a point transformation.*

We may combine these two theorems into a more general one, the proof being analogous to the two used above.

THEOREM 3. *A transformation preserving contact of order m between varieties of dimension $n - r$ in V_n is essentially the $(m-1)^{\text{th}}$ extension of the contact transformation of Lie if $r = 1$, and if $r > 1$ the transformation is essentially the m^{th} extension of a point transformation.*

PRINCETON UNIVERSITY.

ANNALS OF MATHEMATICS
Vol. 39, No. 3, July, 1938

MEROMORPHIC CURVES

BY HERMANN AND JOACHIM WEYL

(Received February 9, 1938)

A meromorphic curve $\mathfrak{C}$ in the k-dimensional projective space R with the homogeneous coördinates $x_0, x_1, \cdots, x_k$ is obtained by setting up $x_1/x_0, \cdots, x_k/x_0$ as meromorphic functions of a complex parameter z ranging over the whole z-plane (excluding $z = \infty$). R. Nevanlinna's theory of meromorphic functions, as the case $k = 1$, may be extended so as to yield a similar theory of meromorphic curves for any number k of dimensions. We follow a simplified method developed by L. Ahlfors.[1]

Sections 1 and 2 are concerned with preliminaries about projective geometry and the local behavior of a meromorphic curve.

1. Distances and Means in Projective Space

In describing $k + 1$ numbers $x_0, x_1, \cdots, x_k$ as components of a *vector* in $(k + 1)$-dimensional vector space, we convey our intention to study only such relations as are invariant under non-singular linear transformations

$$x_i' = \sum_j h_{ij} x_j, \qquad (i, j = 0, 1, \cdots, k), \tag{1.1}$$

in considering them as coördinates of a *point* in projective k-space R we agree moreover not to distinguish between the set x_i and the set ρx_i, ρ being an arbitrary non-vanishing factor. With rare exceptions we shall use the projective terminology.

A *plane* is given by an equation

$$(\alpha x) = \sum_i \alpha_i x_i = 0, \qquad (\alpha_0, \alpha_1, \cdots, \alpha_k) \neq (0, 0, \cdots, 0). \tag{1.2}$$

Its coördinates α_i transform contragrediently to the point coördinates x_i.

l points $x, x', \cdots, x^{(l-1)}$ span a linear $(l - 1)$-*spread* $[xx' \cdots]$ with the coordinates

$$[xx' \cdots]_{i_1 \cdots i_l} = \begin{vmatrix} x_{i_1}, & x_{i_2}, & \cdots, & x_{i_l} \\ x'_{i_1}, & x'_{i_2}, & \cdots, & x'_{i_l} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ x^{(l-1)}_{i_1}, & x^{(l-1)}_{i_2}, & \cdots, & x^{(l-1)}_{i_l} \end{vmatrix}. \tag{1.3}$$

A set $x(l)$ of skew-symmetric numbers $x_{i_1 \cdots i_l}$ which under the influence of (1.1) transform in the same manner as (1.3), is called an $(l - 1)$-*complex* in projective

[1] Soc. Sci. Fenn. Comm. Phys.-Math. **8**, No. 10, 1935.

geometry, where only their ratio is taken into account. The spreads shall be called *special* complexes. The complexes are points in a projective space R_l of

$$k_l = \binom{k+1}{l} - 1$$

dimensions. The linear form in R_l :

$$\sum \alpha_{i_1 \cdots i_l} x_{i_1 \cdots i_l} \qquad (i_1 < i_2 < \cdots < i_l)$$

is denoted as $(\alpha(l), x(l))$ or $(\alpha x)_l$ and in particular as $\alpha(x, x', \cdots)$ for $x(l) = [xx' \cdots x^{(l-1)}]$. Summation indices like $i_1, \cdots, i_l$ are always bound by the restriction $i_1 < i_2 < \cdots < i_l$.

The *distance* $|| \alpha x ||$ between a point x and a plane α may be introduced by

$$|| \alpha x ||^2 = | \sum_i \alpha_i x_i |^2 : \sum_i | \alpha_i |^2 \sum_i | x_i |^2. \tag{1.4}$$

The distance is a real number $\geqq 0$ and $\leqq 1$. It is invariant under *unitary* transformations (1.1). After an arbitrary transformation (1.1), the new distance $|| \alpha x ||'$ differs from the old one $|| \alpha x ||$ by a factor lying between the fixed positive limits Γ/γ and γ/Γ for all α and x. Γ and γ are maximum and minimum of the positive definite Hermitian form

$$\sum_i | x_i' |^2 = \sum_i | \sum_j h_{ij} x_j |^2$$

under the restriction $\sum | x_i |^2 = 1$. This amount of dependence of our unitary metric on the coördinates will prove harmless for the applications we have in mind.

The unitarily invariant *distance* $[[xy]]$ between two points x and y is defined by

$$[[xy]]^2 = \sum_{i<j} | x_i y_j - x_j y_i |^2 : \sum_i | x_i |^2 \sum_i | y_i |^2. \tag{1.5}$$

It is again $\geqq 0$ and $\leqq 1$. The factor it takes on under the influence of (1.1) lies between $(\Gamma/\gamma)^2$ and $(\gamma/\Gamma)^2$.

Similar conditions prevail in R_l.

$$X_l^2 = | x(l) |^2 = \sum | x_{i_1 \cdots i_l} |^2$$

is invariant under unitary transformations (1.1). The numbers Γ_l, γ_l playing the same rôle in R_l as Γ, γ in R are $\leqq \Gamma^l$ and $\geqq \gamma^l$ respectively. The sub-index l always refers to the space R_l.

Let us examine the distance $[[ax]]_l$ of a given point $a(l)$ in R_l from one varying over the whole set S_l of special points

$$x(l) = [xx' \cdots x^{(l-1)}].$$

We maintain that $[[ax]]_l$ takes on a *minimum* $\delta_l(a)$ on S_l. This follows readily from the possibility of spanning any given $(l-1)$-spread $x(l)$ by l *unitary-orthogonal* vectors $x, x', \cdots, x^{(l-1)}$ satisfying the relations

$$\sum_i \bar{x}_i^{(\alpha)} x_i^{(\beta)} = \begin{cases} 1 & (\alpha = \beta) \\ 0 & (\alpha \neq \beta) \end{cases} \qquad (\alpha, \beta = 0, \cdots, l-1).$$

Indeed, not only do these relations define a *compact* manifold in the space of the $(k+1)l$ variables $x_i^{(\alpha)}$ but at the same time, by implying $|x(l)| = 1$, they do away with the denominator in $[[ax]]_l$. $\delta_l(a)$, the distance of a from S_l, will be positive and not zero if a itself is not special. Thus in that case

$$[[ax]]_l \geqq \delta_l(a) > 0 \qquad \text{(for all } x(l) \text{ in } S_l).$$

The unit sphere $\mathfrak{S}_k$,

$$|\alpha_0|^2 + |\alpha_1|^2 + \cdots + |\alpha_k|^2 = 1 \tag{1.6}$$

in the Euclidean space whose coördinates are the real and imaginary parts of $\alpha_0, \alpha_1, \cdots, \alpha_k$, may be called the *unitary k-sphere*. Of any continuous function $\psi(\alpha_0, \alpha_1, \cdots, \alpha_k)$ we can form the *average*

$$\mathfrak{M}_\alpha \psi = \int \psi(\alpha)\, d\omega_\alpha : \int d\omega_\alpha$$

over the whole sphere $\mathfrak{S}_k$, $d\omega_\alpha$ being its volume element at the point α. Because the metric of the sphere is invariant with respect to real orthogonal transformations of the $2(k+1)$ real coördinates and thus *a fortiori* with respect to unitary transformations $\mathfrak{U}$ of $\alpha_0, \cdots, \alpha_k$, our averaging process is invariant under $\mathfrak{U}$.

Two vectors $(\alpha_0, \cdots, \alpha_k)$ and $(\beta_0, \cdots, \beta_k)$ form a *unitary pair* if

$$\sum \bar{\alpha}_i \alpha_i = 1, \qquad \sum \bar{\beta}_i \beta_i = 1, \qquad \sum \bar{\alpha}_i \beta_i = 0. \tag{1.7}$$

The locus of the unitary pairs is a manifold $\mathfrak{S}\mathfrak{S}_k$ of $4k$ real dimensions in the $4(k+1)$-dimensional Euclidean space with the real and imaginary parts of α_i, β_i as its coördinates and

$$|d\alpha_0|^2 + \cdots + |d\alpha_k|^2 + |d\beta_0|^2 + \cdots + |d\beta_k|^2$$

as its metrical ground form. $\mathfrak{S}\mathfrak{S}_k$ is invariant under unitary transformations $\mathfrak{U}$ cogrediently applied to α and β and therefore homogeneous. As the study of the infinitesimal neighborhood of

$$\alpha = (1, 0, 0, \cdots, 0), \qquad \beta = (0, 1, 0, \cdots, 0)$$

at once reveals, it is therefore a differentiable manifold free of singularities and thus, by virtue of the imbedding Euclidean space endowed with a non-degenerate metric, invariant under $\mathfrak{U}$. Its volume element $d\omega_{\alpha,\beta}$ allows us to define an averaging process $\mathfrak{M}_{\alpha\beta}$ extending over $\mathfrak{S}\mathfrak{S}_k$ which has the same property of invariance. If the process is applied to a function $\psi(\alpha)$ depending only on

α and not on β, we maintain that its effect will be the same as by taking the average of $\psi(\alpha)$ with respect to α alone ranging over $\mathfrak{S}_k$. We state this fact as

LEMMA 1. $\mathfrak{M}_{\alpha\beta}(\psi(\alpha)) = \mathfrak{M}_{\alpha}(\psi(\alpha))$.

Indeed, the set of all points (α, β) on $\mathfrak{S}\mathfrak{S}_k$ for which α lies in the given element $d\omega_\alpha$ will have a certain volume $G(\alpha)\cdot d\omega_\alpha$. From the unitary invariance it is evident that the function $G(\alpha)$ of α is actually a constant G independent of α. This observation proves our lemma.

We shall have to use the mean value $\mathfrak{M}_{\alpha}$ of the particular function

$$\log \frac{1}{\| \alpha x \|}. \tag{1.8}$$

LEMMA 2.

$$\mathfrak{M}_{\alpha} \log \frac{1}{\| \alpha x \|}$$

is a constant m^0 independent of $(x_0, \cdots, x_k)$.

This is an immediate consequence of the unitary invariance of $\| \alpha x \|$ and of the process $\mathfrak{M}_{\alpha}$. Therefore our mean has the same value as for $x = (1, 0, \cdots, 0)$ or

$$\mathfrak{M}_{\alpha} \log \frac{1}{|\alpha_0|} = m^0.$$

Incidentally one easily computes

$$m^0 = \frac{1}{2}\left(\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{k}\right).$$

Our result may be stated in the form:

$$\mathfrak{M}_{\alpha} \log |\alpha_0 x_0 + \cdots + \alpha_k x_k| = \log |x| + \text{const.}$$

This remark that the averaging process changes the modulus of the linear form $\alpha_0 x_0 + \cdots + \alpha_k x_k$ under the logarithm into the unitary length $|x|$,

$$|x|^2 = |x_0|^2 + \cdots + |x_k|^2,$$

will be of great importance later on.

Let us further compute the mean $\mathfrak{M}^a_{\alpha}$ of the same function (1.8) over the unitary $(k-1)$-sphere $\mathfrak{S}_{k-1}(a)$ cut out from $\mathfrak{S}_k$ by a plane:

$$a_0\alpha_0 + \cdots + a_k\alpha_k = 0, \qquad |\alpha_0|^2 + \cdots + |\alpha_k|^2 = 1. \tag{1.9}$$

In returning to our former dual standpoint where α is taken as a plane rather than a point, one will describe this process as averaging over all planes α through a given point a.

LEMMA 3. $\mathfrak{M}_{a}^{a} \log \frac{1}{\|\alpha x\|} = \log \frac{1}{[[ax]]} + \text{const.}$

PROOF. On account of unitary invariance one may assume

$$a_0 = 1, \qquad a_1 = \cdots = a_k = 0. \tag{1.10}$$

We then have $\alpha_0 = 0$ on $\mathfrak{S}_{k-1}(a)$, $\mathfrak{M}_{\alpha}^{a}$ is the mean over the sphere $\mathfrak{S}_{k-1}$,

$$|\alpha_1|^2 + \cdots + |\alpha_k|^2 = 1,$$

and on $\mathfrak{S}_{k-1}$

$$\log \frac{1}{\|\alpha x\|^2} = \log \frac{(|\alpha_1|^2 + \cdots + |\alpha_k|^2)(|x_1|^2 + \cdots + |x_k|^2)}{|\alpha_1 x_1 + \cdots + \alpha_k x_k|^2} + \log \frac{|x_0|^2 + |x_1|^2 + \cdots + |x_k|^2}{|x_1|^2 + \cdots + |x_k|^2}.$$

By Lemma 2, applied to $k - 1$ instead of k, the mean of the first part is a constant. The second part, which is independent of α, equals

$$\log \frac{1}{[[ax]]^2}$$

for (1.10).

2. Local Investigation of a Meromorphic Curve

In order to accomplish complete symmetry in all coördinates we prefer to define a *meromorphic curve* $\mathfrak{C}$ as follows. To each value $z = z_0$ of the parameter z there are ascribed $k + 1$ function elements

$$x_i = x_i(z) = c_i + c_i'(z - z_0) + \cdots \qquad (i = 0, 1, \cdots, k) \tag{2.1}$$

such that

$$(c_0, c_1, \cdots, c_k) \neq (0, 0, \cdots, 0). \tag{2.2}$$

We are free to replace $x_i(z)$ by $\rho(z)x_i(z)$ with a common factor

$$\rho(z) = \rho_0 + \rho_1(z - z_0) + \cdots \qquad (\rho_0 \neq 0) \tag{2.3}$$

not vanishing at the center z_0. $(c_0 : \cdots : c_k)$ is the point $z = z_0$ of the curve. The elements corresponding to a value z_1 sufficiently near to z_0 are obtained from (2.1) by rearranging the power series according to powers of $z - z_1$ (direct analytic continuation). *We assume that the curve lies in no linear subspace of R.*

A given linear form with constant coefficients α_i,

$$(\alpha x) = \alpha_0 \cdot x_0(z) + \cdots + \alpha_k \cdot x_k(z), \qquad (\alpha_0, \cdots, \alpha_k) \neq (0, \cdots, 0),$$

will vanish to a certain finite order $h = h(z_0; \alpha) \geqq 0$ at $z = z_0$. h is the multiplicity in which the plane α cuts the curve $\mathfrak{C}$ at z_0; the multiplicity obviously is not affected by the arbitrary gauge factor (2.3).

The curve may be looked upon not only as the set of its points, but also as the set of its tangents (1-spreads), its osculating planes (2-spreads), and so on. By means of the elements (2.1) and *their successive derivatives* $x'(z)$, $x''(z)$, $\cdots$ we form $[xx' \cdots x^{(l-1)}]$ with the components

$$x_{i_1 \cdots i_l} = \begin{vmatrix} x_{i_1}(z), & \cdots, & x_{i_l}(z) \\ x'_{i_1}(z), & \cdots, & x'_{i_l}(z) \\ \cdots\cdots & & \cdots\cdots \end{vmatrix} \tag{2.4}$$

and thus obtain a meromorphic curve $\mathfrak{C}_l$ in R_l : this is "the curve $\mathfrak{C}$ considered as the set of its $(l-1)$-spreads." The definition is based on the fundamental fact that a change of gauge $x_i(z) \to \rho(z)x_i(z)$ has upon the determinants (2.4) the effect of multiplying them all with the common factor ρ^l. l is any of the numbers $1, \cdots, k+1$. However one has to be careful in that, contrary to the convention (2.2), all the elements (2.4) might simultaneously vanish at z_0 by containing some common power $(z - z_0)^{d_l}$. Only after removing this power do we obtain those elements

$$x_{i_1 \cdots i_l}(z) = c_{i_1 \cdots i_l} + c'_{i_1 \cdots i_l}(z - z_0) + \cdots \tag{2.5}$$

which define our curve $\mathfrak{C}_l$ in the vicinity of z_0 and whose initial coefficients $c_{i_1 \cdots i_l}$ do not all vanish. We reserve the notation $x_{i_1 \cdots i_l}(z)$ for these expansions (2.5) while the determinants (2.4) are denoted by $[xx' \cdots]_{i_1 \cdots i_l}$; we have

$$[x(z), x'(z), \cdots]_{i_1 \cdots i_l} = (z - z_0)^{d_l} \cdot x_{i_1 \cdots i_l}(z).$$

When we are to determine the multiplicity $h_l = h_l(z_0 ; \alpha)$ of the intersection of a "plane"

$$\sum \alpha_{i_1 \cdots i_l} x_{i_1 \cdots i_l} = 0 \tag{2.6}$$

in R_l with $\mathfrak{C}_l$ at the point z_0 we substitute the expansions (2.5) in the left side of (2.6). In other words, the *fixed* zero of order d_l which

$$\alpha(x(z), x'(z), \cdots) = \sum \alpha_{i_1 \cdots i_l}[x(z), x'(z), \cdots]_{i_1 \cdots i_l}$$

possesses at $z = z_0$ independent of the values of the coefficients α *does not count.*

According to a well-known method, the local investigation is carried out as follows. Sticking to the hypothesis that $\mathfrak{C}$ lies in no linear subspace of less dimensions, we can ascertain a suitable projective coördinate system such that the expansions

$$x_0 = 1 + \cdots, \qquad x_1 = (z - z_0)^{f_1} + \cdots, \cdots, \qquad x_k = (z - z_0)^{f_k} + \cdots \tag{2.7}$$

start with terms of ever higher order:

$$0 = f_0 < f_1 < \cdots < f_k .$$

We use the differences

$$e_0 = 0, \qquad e_1 = f_1 - f_0 , \qquad e_2 = f_2 - f_1 , \cdots , \qquad e_k = f_k - f_{k-1} .$$

The point $z = z_0$ of the curve $\mathfrak{C}$ is then $(1 : 0 : \cdots : 0)$. A plane $(\alpha_0, \alpha_1, \cdots, \alpha_k)$ with $\alpha_0 \neq 0$ does not pass through this point and hence its multiplicity of intersection $h(z_0 ; \alpha) = 0$. But all planes *through the point* cut the curve at z_0 *at least* $f_1 = e_1$ *times* and in general (i.e. unless the α satisfy further linear conditions) not more often. $e_1 > 1$ indicates that the point of the curve is stationary or a cusp. Hence we call $e_1 - 1 = e_1(z_0) - 1$ the *local stationarity index*. In the same manner we study $\mathfrak{C}_l$ at z_0. The expansion of the determinant

$$(i_1 \cdots i_l) = [x(z), x'(z), \cdots]_{i_1 \cdots i_l} \tag{2.8}$$

begins exactly with the power

$$(f_{i_1} - 0) + (f_{i_2} - 1) + \cdots + (f_{i_l} - l + 1)$$

of $z - z_0$. Hence

$$\begin{aligned} d_l &= f_0 + (f_1 - 1) + \cdots + (f_{l-1} - l + 1) \\ &= (l - 1)(e_1 - 1) + (l - 2)(e_2 - 1) + \cdots + 1(e_{l-1} - 1), \end{aligned}$$

and the point $z = z_0$ of the curve $\mathfrak{C}_l$ has the coördinates:

$$c_{0,1,\cdots,l-1} = 1, \text{ all other } c_{i_1 \cdots i_l} = 0.$$

Incidentally this shows that the point is in S_l. Any plane (2.6) which cuts our curve $\mathfrak{C}_l$ at all in $z = z_0$, $\alpha_{0,1,\cdots,l-1} = 0$, cuts it there at least e_l times and in general not more often, because all (2.8) except $(0, 1, \cdots, l - 1)$ vanish to the order

$$f_0 + (f_1 - 1) + \cdots + (f_{l-2} - l + 2) + (f_l - l + 1) = d_l + e_l$$

at least, and $(0, \cdots, l - 2, l)$ exactly to this order. Hence $e_l - 1$ is the stationarity index of $\mathfrak{C}_l$. (A stationary tangent is called an inflection tangent; everyday language provides no terms for the higher l.) This argument shows that e_l plays a two-fold rôle: on the one hand the $e_l - 1$ are the second differences of the fixed zero orders,

$$e_l - 1 = d_{l+1} - 2d_l + d_{l-1} \qquad (l = 1, \cdots, k),$$

on the other hand $e_l - 1$ is *the* l^{th} *stationarity index.*

3. The First Main Theorem

A plane α cuts the curve $\mathfrak{C}$ in

$$n(r; \alpha) = \sum_{|z|<r} h(z; \alpha)$$

points within the z-circle of radius r. Let α, β be two planes. Application of the principle of the argument to the meromorphic function

$$w = (\alpha x) : (\beta x) = (\alpha_0 x_0 + \cdots + \alpha_k x_k) : (\beta_0 x_0 + \cdots + \beta_k x_k) \tag{3.1}$$

results in the equation

$$n(r;\alpha) - n(r;\beta) = r\frac{d}{dr}\left\{\frac{1}{2\pi}\int_0^{2\pi}\log\frac{|(\alpha x)|}{|(\beta x)|}\,d\varphi\right\}. \tag{3.2}$$

Integration by φ always refers to the z-circle of radius r, $z = re^{i\varphi}$, and extends over the whole circumference from $\varphi = 0$ to 2π. We should like to split the log on the right side into the two terms

$$\log|(\alpha x)| - \log|(\beta x)|.$$

This will not do, however, in view of the fact that only the *ratios* of the α_i and of the x_i have a meaning. Ahlfors's trick by which he overcomes this obstacle (in the case $k = 1$) consists in *replacing* $|(\alpha x)|$ *by the distance* $\|\alpha x\|$. We are thus prompted to introduce the "*compensating function*"

$$m(r;\alpha) = \frac{1}{2\pi}\int_0^{2\pi}\log\frac{1}{\|\alpha x\|}\cdot d\varphi. \tag{3.3}$$

(3.2) now states that

$$n(r;\alpha) + r\frac{d}{dr}m(r;\alpha) = t(r)$$

is independent of the plane α. When, after division by r, we integrate from a lower limit $r_0 > 0$ fixed once for all, we find

$$N(r;\alpha) + m(r;\alpha) = T(r) + m(r_0;\alpha) \tag{3.4}$$

with

$$N(r;\alpha) = \int_{r_0}^{r} n(r;\alpha)\frac{dr}{r}, \qquad T(r) = \int_{r_0}^{r} t(r)\frac{dr}{r}.$$

A similar notation shall always be employed: for any function $t(r)$ we denote the integral of $t(r)/r$ from r_0 to r by the corresponding capital $T(r)$.

In the relation (3.4) it is essential that the compensating term $m(r;\alpha) \geqq 0$ and that $T(r)$ is independent of α. The constant term $m(r_0;\alpha)$ is of relatively slight importance. We notice the inequality

$$N(r;\alpha) \leqq T(r) + m(r_0;\alpha). \tag{3.5}$$

The function N vanishes for $r = r_0$ and is an *increasing convex function of* $\log r$. Any function of r with these properties shall be called *of regular type*. For a given value of r the function $m(r;\alpha)$ is continuous in α except at the origin $(\alpha_0, \cdots, \alpha_k) = (0, \cdots, 0)$; but being homogeneous it is even bounded around the origin. According to (3.4) the same conditions prevail for $N(r;\alpha)$.

Replacement of $|(\alpha x)|$ by $\|\alpha x\|$ has the disadvantage of introducing into our projective space a *unitary metric* dependent on the coördinate system. However, as we mentioned, a change of coördinates modifies

$$\log\frac{1}{\|\alpha x\|} \text{ and hence } m(r;\alpha)$$

merely by an additive term which remains in absolute value $\leqq \log \Gamma - \log \gamma$. Let us agree to call two functions of r *equivalent* ($\sim$) if their difference remains bounded as r approaches infinity. We then find that the function $T(r)$ is practically, i.e. in the sense of equivalence, independent of the projective coordinate system employed. The essential content of the equation (3.4) is better emphasized by writing it as an equivalence:

$$N(r; \alpha) + m(r; \alpha) \sim T(r). \tag{3.6}$$

In the rational case, where one deals with a rational curve of order n, one has $T(r) \sim n \log r$; for then

$$N(r; \alpha) \sim n'(\alpha) \log r, \qquad m(r; \alpha) = n_\infty(\alpha) \log r,$$

if $n_\infty(\alpha)$, $n'(\alpha)$ designate the numbers of intersecting points of the plane α at $z = \infty$ and for all finite z respectively. It was in view of this situation that we called m the compensating function: in the meromorphic case it makes up for the lacking contribution from $z = \infty$. In studying meromorphic functions we have no right to consider the infinite of the z-plane as a point; it is rather an infinitely large circle. The formula (3.6) is the analogue of the theorem that every plane cuts a rational curve at the same number of points. Therefore we call the function $T(r)$ the *order* of the meromorphic curve $\mathfrak{C}$. (Two curves are *of the same order* if their functions $T(r)$ are *equivalent*.)

This analysis clearly indicates that in the *defect relations* one must expect not the function $m(r; \alpha)$ but its *mean over all planes α through a given point a* to play the rôle of the defect. We are thus induced to try two things: to average the fundamental equation (3.4) first over *all* planes α and then over all planes *through a*. Lemma 2 tells us that the average

$$\mathfrak{M}_\alpha\, m(r; \alpha) = m(r)$$

is a constant m^0, independent of r. This result is important enough in itself; for it reveals the surprising fact that *in the average the compensating term is ~ 0* and therefore completely negligible. Together with (3.4) we obtain the equation

$$T(r) = N(r) \tag{3.7}$$

where $N(r)$ arises in the usual way by integration from

$$n(r) = \mathfrak{M}_\alpha\, n(r; \alpha).$$

Consequently *the order $T(r)$ itself is a function of regular type.*

When averaging over the planes α through a given point a one will introduce

$$n^*(r; a) = \mathfrak{M}^a_\alpha\, n(r; \alpha)$$

and derive from Lemma 3 the relation

$$\mathfrak{M}^a_\alpha\, m(r; \alpha) = m^*(r; a) + \text{const.}$$

with

$$m^*(r;a) = \frac{1}{2\pi}\int_0^{2\pi} \log \frac{1}{[[ax]]}\, d\varphi. \tag{3.8}$$

Thus (3.4) is paralleled by

$$N^*(r;a) + m^*(r;a) = T(r) + m^*(r_0;a). \tag{3.9}$$

We may apply all this to the curve $\mathfrak{C}_l$ instead of $\mathfrak{C}$. The resulting equations

$$\boxed{\begin{aligned} N_l(r;\alpha) + m_l(r;\alpha) &= T_l(r) + m_l(r_0;\alpha), \\ N_l(r) &= T_l(r), \\ N_l^*(r;a) + m_l^*(r;a) &= T_l(r) + m_l^*(r_0;a) \end{aligned}} \qquad (l = 1, \cdots, k) \tag{3.10}$$

constitute the *first main theorem*. *All N's are functions of regular type, the m's are* $\geqq 0$. The N and m and T are practically independent of the coördinate system although they are strictly invariant only under unitary transformations. Whenever a is not special,

$$m_l^*(r;a) \leqq \log 1/\delta_l(a), \qquad m_l^*(r;a) \sim 0.$$

$\mathfrak{C}_l$ might well happen to lie in a linear subspace of R_l of dimensionality $< k_l$. In order to apply to $\mathfrak{C}_l$ our theory developed for $\mathfrak{C}$ we must modify the latter so as to cover a *degenerate* $\mathfrak{C}$ lying in a subspace of R. The little changes needed in that case and their influence upon our further arguments will be discussed in the appendix. For the time being we treat $\mathfrak{C}_l$ as a non-degenerate curve in R_l.

4. Second Main Theorem

We introduce the numbers

$$v_l(r) = \sum_{|z|<r} \{e_l(z) - 1\}$$

of stationary $(l-1)$-spreads of our curve $\mathfrak{C}$ within the circle $|z| < r$. The $v_l(r)$ are the second differences of the

$$u_l(r) = \sum_{|z|<r} d_l(z).$$

With arbitrary coefficients $\alpha_{i_1 \cdots i_l}$, β_i we form the meromorphic function

$$\zeta = (\alpha x)_l : (\beta x)^l = \alpha(x(z), \cdots, x^{(l-1)}(z)) : \{\beta(x(z))\}^l$$

and apply the principle of the argument:

$$u_l(r) + n_l(r;\alpha) - ln(r;\beta) = r\frac{d}{dr}\left\{\frac{1}{2\pi}\int_0^{2\pi} \log|\zeta|\, d\varphi\right\},$$

$$U_l(r) + N_l(r;\alpha) - lN(r;\beta) = \frac{1}{2\pi}\int_0^{2\pi} \log|\zeta|\, d\varphi + \text{const.}$$

Add $m_l(r;\alpha) - lm(r;\beta)$ on both sides. On the left we shall get $U_l(r) + T_l(r) - lT(r) + \text{const.}$, on the right

$$\frac{1}{2\pi}\int_0^{2\pi} \log\,(X_l : X^l)\,d\varphi + \text{const.},$$

where

$$X_l^2 = \sum |\,[x(z), \cdots, x^{(l-1)}(z)]_{i_1 \cdots i_l}\,|^2, \qquad X^2 = \sum |\,x_i(z)\,|^2.$$

Under the influence of the gauge factor $\rho = \rho(z)$ both numerator and denominator of $X_l : X^l$ take on the same factor $|\,\rho\,|^l$; the fraction itself stays unaltered. The V_l being the second differences of the U_l, we find it better to pass to the second differences of the equations just derived:

$$\boxed{V_l(r) + \{T_{l+1}(r) - 2T_l(r) + T_{l-1}(r)\} = \Omega_l(r) + C_l} \qquad (l = 1, \cdots, k). \tag{4.1}$$

C_l is a constant and

$$\Omega_l(r) = \frac{1}{2\pi}\int_0^{2\pi} \log\,(X_{l+1}X_{l-1} : X_l^2)\,d\varphi. \tag{4.2}$$

For the extreme cases $l = 1$ and k one ought to take account of the conventions

$$X_0 = 1; \qquad T_0(r) = T_{k+1}(r) = 0.$$

The relations (4.1) which link *the l-indices* $V_l(r)$ to the *l*-orders $T_l(r)$ constitute the *second main theorem.* *$V_l(r)$ is a function of regular type.*

(4.1) are the analogues of the equations

$$v_l + (n_{l+1} - 2n_l + n_{l-1}) = -2,$$

holding in the rational case between the orders n_l and the total numbers v_l of stationary $(l-1)$-spreads. Indeed, if e_l^∞ denote the exponents similar to e_l in the expansions (2.7) for the neighborhood of $z = \infty$, progressing by powers of $1/z$, then $e_l^\infty - 1$ are the local indices for that point, while one readily finds

$$\log\,(X_{l+1}X_{l-1} : X_l^2) \sim -(e_l^\infty + 1)\log r. \tag{4.3}$$

$V_l(r)$ is $\sim v_l' \log r$, where v_l' is the sum $\sum_z \{e_l(z) - 1\}$ extending over the finite values z only. Here the $e_l - 1$ appear throughout as the second differences of the d_l.

Their other meaning, as described by the word stationarity index, or by the equation

$$e_l(\mathfrak{G}) = e_1(\mathfrak{G}_l) \tag{4.4}$$

could be exhibited in this way. In R_l we form, by means of the expansions (2.5), the integral

$$\frac{1}{2\pi}\int_0^{2\pi} \log \frac{1}{[[ax]]_l}\,d\varphi$$

extending over a small circle $z - z_0 = r \cdot e^{i\varphi}$ around the center z_0. With r tending to zero, the integral will stay finite (be ~ 0) unless a coincides with the point c, $z = z_0$, of the curve $\mathfrak{C}_l$. In the latter case it is $\sim e_l \log 1/r$. By applying this to the point $z = \infty$ of a rational curve we find

$$m_l^*(r; a) \sim e_l^\infty \log r \text{ or } \sim 0 \tag{4.5}$$

according as a is the point c, $z = \infty$, of the curve or not. Since we are not allowed to consider $z = \infty$ as a single point c in the meromorphic case, (4.3) and (4.5) suggest the following meromorphic generalization,

$$\Omega_l(r) + \sum_a m_l^*(r; a) = -\log r + \cdots, \tag{4.6}$$

where the sum extends over any number q of $(l-1)$-complexes $a = a^{(1)}, \cdots, a^{(q)}$ and $\cdots$ indicates an "essentially" negative term. The proof of this equation, which, as it were, is the analogue of the relation (4.4) for the point $z = \infty$, must bring out what we mean exactly by "essentially negative." For the validity of (4.6), the hypothesis that $a^{(1)}, \cdots, a^{(q)}$ are *distinct* is certainly relevant. Beyond that, we shall be forced to suppose $a^{(1)}, \cdots, a^{(q)}$ to *satisfy no accidental linear relation*. (An accidental linear relation for a given finite set of points $a^{(1)}, \cdots, a^{(q)}$ in a k-dimensional projective space is a relation between any $p \leqq k+1$ of those points, $a^{(1)}, \cdots, a^{(p)}$ say,

$$\lambda_1 a^{(1)} + \cdots + \lambda_p a^{(p)} = 0$$

with non-vanishing numerical coefficients λ. Linear relations between more than $k+1$ points are inevitable.)

In §5 we shall deduce the estimate

$$\Omega_l(r) = -\log r + \cdots;$$

in §6 we add the sum of defects m^* as in (4.6).

5. Estimate of the Compensating Term Ω_l

Let us first observe that

$$T(r) \geqq \log r - \text{const.}, \tag{5.1}$$

as follows at once from (3.4) by choosing for α a plane passing through the point $z = 0$ of the curve (or through any other point). The same inequality holds for $T_l(r)$.

In our investigation we shall encounter functions $\theta(r)$ such that $\lambda^*(r) = \exp(\theta(r))$ gives rise by twofold integration

$$\lambda(r) = \int_0^r \lambda^*(r) r\, dr, \qquad \Lambda(r) = \int_{r_0}^r \lambda(r) \frac{dr}{r}$$

to a function

$$\Lambda(r) \leqq cT(r) + c' \qquad (c, c' \text{ any two constants}).$$

This fact shall be expressed by

$$\theta(r) = \omega(T). \tag{5.2}$$

If $\theta = \omega(T)$ then the same holds for any function $\leqq \theta(r)$ and for $\theta(r) + \text{const.}$ The implications of (5.2) may be exhibited more explicitly in various ways; we shall be content with one simple consequence. Let κ now and forever be any fixed number > 1. From

$$dr/r = d\Lambda/\lambda$$

it follows that the integral of dr/r is finite when extended over those intervals where

$$1/\lambda < 1/\Lambda^{\kappa}.$$

Outside these intervals we have $\lambda \leqq \Lambda^{\kappa}$. In the same manner we find that "*almost everywhere*", i.e. with the exception of a set of intervals over which the integral $\int dr/r$ stays finite, the inequality

$$\lambda^* \leqq r^{-2}\lambda^{\kappa}$$

prevails. Hence almost everywhere

$$\theta = \log \lambda^* \leqq \kappa^2 \log \Lambda - 2 \log r.$$

Here κ^2 is like κ itself any number > 1; therefore we may replace it by κ. Taking (5.1) into account, we see that (5.2) implies the inequality

$$\theta(r) \leqq \kappa \log T(r) - 2 \log r$$

almost everywhere.

Again we study the meromorphic function (3.1),

$$(\alpha x) : (\beta x) = w(z) = w = w_1/w_2 .$$

We plot its values w on the w-sphere of diameter 1 into which the w-plane changes by stereographic projection. Let

$$d\tau_w = dw\, d\bar{w}/(1 + w\bar{w})^2$$

denote its surface element (spanned by the two line elements dw and $\sqrt{-1}\,dw$ at the point w). The function $w(z)$ maps the interior of the circle $|z| < r$ in a one-to-one fashion upon a Riemann surface covering the w-sphere. We evaluate its area $2\pi f(r; \alpha, \beta)$ in two ways. First we have

$$f(r; \alpha, \beta) = \frac{1}{2\pi} \int_0^r \int_0^{2\pi} \frac{|w'(z)|^2}{(1 + w\bar{w})^2}\, d\varphi \cdot r\, dr; \tag{5.3}$$

on the other hand, since the point w is covered by the Riemann surface

$$n(r; \alpha_i w_2 - \beta_i w_1)$$

times,

$$f(r;\alpha,\beta)=\frac{1}{2\pi}\int n(r;\alpha_i w_2-\beta_i w_1)\,d\tau_w.$$

For the derivative $w'(z)$ we find

$$w'(z)=\sum_{i<j}[\beta\alpha]_{ij}[xx']_{ij}:\Big(\sum_i \beta_i x_i\Big)^2.$$

By averaging with respect to α and β over the manifold $\mathfrak{S}\mathfrak{S}_k$ described in §1, we obtain on the one hand for

$$\lambda(r)=\mathop{\mathfrak{M}}_{\alpha\beta} f(r;\alpha,\beta)$$

the equation

$$\lambda(r)=\int_0^r \lambda^*(r)r\,dr$$

with

$$\lambda^*(r)=\frac{1}{2\pi}\int_0^{2\pi}\mathop{\mathfrak{M}}_{\alpha\beta}(Q^2)\,d\varphi,$$

(5.4)
$$Q=\frac{\Big|\sum_{i<j}[\alpha\beta]_{ij}[xx']_{ij}\Big|}{|\sum\alpha_i x_i|^2+|\sum\beta_i x_i|^2},$$

on the other hand for

$$\Lambda(r)=\int_{r_0}^r \lambda(r)\frac{dr}{r}$$

the relation

$$\Lambda(r)=\frac{1}{2\pi}\int \mathop{\mathfrak{M}}_{\alpha\beta} N(r;\alpha_i w_2-\beta_i w_1)\,d\tau_w.$$

When writing w in homogeneous form $=w_1/w_2$ we find it convenient to normalize by

$$w_1\overline{w}_1+w_2\overline{w}_2=1.$$

For given w_1, w_2 the vectors

(5.5)
$$\xi_i=\alpha_i w_2-\beta_i w_1,\qquad \eta_i=\alpha_i\overline{w}_1+\beta_i\overline{w}_2$$

form a unitary pair just like α and β, and in particular

(5.6)
$$|\xi_0|^2+\cdots+|\xi_k|^2=w_1\overline{w}_1+w_2\overline{w}_2=1.$$

Faced with the task of integrating a function $\psi(\xi_0,\cdots,\xi_k)$ with respect to α,β over $\mathfrak{S}\mathfrak{S}_k$ we may first substitute the variables ξ, η for α, β and shall then realize by Lemma 1 that

(5.7)
$$\mathop{\mathfrak{M}}_{\alpha,\beta}\psi(\xi)=\mathop{\mathfrak{M}}_{\xi,\eta}\psi(\xi)=\mathop{\mathfrak{M}}_{\xi}\psi(\xi),$$

and even

$$\frac{1}{2\pi}\int \mathop{\mathfrak{M}}_{\alpha,\beta}\psi(\xi)\cdot d\tau_w = \frac{1}{2}\mathop{\mathfrak{M}}_{\xi}\psi(\xi).$$

Hence

$$\Lambda(r) = \tfrac{1}{2}\mathop{\mathfrak{M}}_{\xi}N(r;\xi) = \tfrac{1}{2}N(r) = \tfrac{1}{2}T(r). \tag{5.8}$$

The logarithm is a *concave* function. Therefore for any positive quantities y and their average $\mathfrak{M}y$ with assigned positive weights, we have

$$\mathfrak{M}\,(\log y) \leqq \log\,(\mathfrak{M}y).$$

Let us apply this to (5.4) with $y = Q^2$ and the average

$$\mathfrak{M} = \frac{1}{2\pi}\int_0^{2\pi} \mathop{\mathfrak{M}}_{\alpha\beta}\cdots d\varphi.$$

We get

$$\frac{1}{2\pi}\int_0^{2\pi} \mathop{\mathfrak{M}}_{\alpha\beta}\log Q^2\cdot d\varphi \leqq \log\lambda^* = \omega(T). \tag{5.9}$$

The problem of computing the mean $\mathop{\mathfrak{M}}_{\alpha\beta}\log Q$ for any two points x, x' has nothing to do with the curve and is fairly easily solved by the remark in §1 that our means change the absolute values of linear forms under a logarithm into the corresponding "lengths." Thus we shall derive the equation

$$\mathop{\mathfrak{M}}_{\alpha,\beta}\log Q = -A + \log\,(X_2 : X_1^2), \qquad A = \text{const.} \tag{5.10}$$

Indeed, on account of unitary invariance we may assume

$$x = (x_0\,,\ 0,\ \ 0,\ \cdots,\ 0),$$
$$x' = (x_0'\,,\ x_1'\,,\ 0,\ \cdots,\ 0).$$

Then $\log Q$ becomes

$$\log\frac{|\alpha_0\beta_1 - \alpha_1\beta_0|}{|\alpha_0|^2 + |\beta_0|^2} + \log\frac{|x_0x_1'|}{|x_0|^2},$$

and we find (5.10) confirmed with

$$A = A(k) = -\mathop{\mathfrak{M}}_{\alpha\beta}\log\frac{|\alpha_0\beta_1 - \alpha_1\beta_0|}{|\alpha_0|^2 + |\beta_0|^2}$$
$$= -\mathop{\mathfrak{M}}_{\alpha\beta}\log\frac{|\alpha_0\beta_1 - \alpha_1\beta_0|}{\sqrt{|\alpha_0|^2 + |\beta_0|^2}\cdot\sqrt{|\alpha_1|^2 + |\beta_1|^2}}.$$

The result is

$$2\Omega_1(r) = \frac{1}{\pi}\int_0^{2\pi}\log\,(X_2 : X_1^2)\,d\varphi = \omega(T),$$

in particular

$$\Omega_1(r) < \frac{\kappa}{2} \log T(r) - \log r$$

"almost everywhere."

In the lowest case $k = 1$ our Q is independent of α and β, and taking the mean becomes superfluous. This is Ahlfors's procedure. The introduction of α and β and the averaging process $\mathfrak{M}_{\alpha\beta}$ are, however, inevitable for higher k; it is the essential new point in our treatment.

We strike against a little formal obstacle when we try to carry our argument over from $l = 1$ to higher l. We first obtain

$$Q = |\, \alpha(x)\beta^*(x) - \alpha^*(x)\beta(x) \,| : (\,|\, \alpha(x) \,|^2 + |\, \beta(x) \,|^2)$$

with the abbreviations

$$\alpha(x) = \alpha(x, x', \cdots, x^{(l-2)}, x^{(l-1)}), \quad \alpha^*(x) = \alpha(x, x', \cdots, x^{(l-2)}, x^{(l)}),$$

and then argue as follows. A recursive substitution with numerical coefficients ρ,

$$\begin{aligned} x &= \rho_0 y, \\ x' &= \rho_{10} y + \rho_1 y', \\ x'' &= \rho_{20} y + \rho_{21} y' + \rho_2 y'', \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x^{(l)} &= \rho_{l0} y + \cdots + \rho_{l,l-1} y^{(l-1)} + \rho_l y^{(l)} \end{aligned}$$

changes $\alpha(x)$, $\alpha^*(x)$ and hence $Q(x)$ according to the equations

$$\alpha(x) = \rho_0 \cdots \rho_{l-1} \cdot \alpha(y), \qquad \alpha^*(x) = \rho_0 \cdots \rho_{l-2}(\rho_l \alpha^*(y) + \rho_{l,l-1}\alpha(y)),$$

$$Q(x) = Q(y) \,|\, \rho_l : \rho_{l-1} \,|\,.$$

Consequently for points of the particular form

$$\begin{aligned} x &= (x_0, 0, 0, \cdots, 0), \\ x' &= (x_0', x_1', 0, \cdots, 0), \\ x'' &= (x_0'', x_1'', x_2'', \cdots, 0), \\ &\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned}$$

$Q(x)$ becomes equal to $|\, x_l^{(l)} \,| : |\, x_{l-1}^{(l-1)} \,|$ times its value

$$Q_0 = \frac{|\, \alpha_0\beta_1 - \alpha_1\beta_0 \,|}{|\, \alpha_0 \,|^2 + |\, \beta_0 \,|^2}$$

for the unit vectors

$$\begin{aligned} x &= (1, 0, 0, \cdots, 0), \\ x' &= (0, 1, 0, \cdots, 0), \\ x'' &= (0, 0, 1, \cdots, 0), \\ &\cdots\cdots\cdots\cdots\cdots \end{aligned}$$

α_0, α_1 are the components with the indices $0, \cdots, l-2, l-1$ and $0, \cdots, l-2, l$.

The road is now open which leads to the relation paralleling (5.10),

$$\mathop{\mathfrak{M}}_{\alpha\beta} \log Q = -A_l + \log (X_{l+1}X_{l-1} : X_l^2) \tag{5.11}$$

with

$$A_l = -\mathop{\mathfrak{M}}_{\alpha\beta} \log \frac{|\alpha_0\beta_1 - \alpha_1\beta_0|}{|\alpha_0|^2 + |\beta_0|^2} = A(k_l).$$

In the latter expression α_0, α_1 stand for any two of the $k_l + 1$ components $\alpha_{i_1 \cdots i_l}$. We find again

$$2\Omega_l(r) = \omega(T_l),$$

and in particular

$$\Omega_l(r) < \frac{\kappa}{2} \log T_l(r) - \log r$$

almost everywhere.

All inequalities in the following paragraph are understood not in the strict sense, but as holding almost everywhere. Omitting in our last relation the negative term $-\log r$, which carries but little weight except in the rational and the lowest transcendental cases, we are for the time being content with the rougher estimate

$$\Omega_l(r) < \frac{\kappa}{2} \log T_l(r). \tag{5.12}$$

From it flow a number of valuable items of information about the *growth of the orders and indices*. First we show that T_l in (5.12) may be replaced by $T = T_1$. Indeed, let us suppose we are in possession of the inequalities

$$\Omega_1 < \frac{\kappa}{2} \log T, \cdots, \Omega_{l-1} < \frac{\kappa}{2} \log T. \tag{5.13}$$

We resort to our old equation

$$\begin{aligned} &\{(l-1)\, V_1 + \cdots + 1\, V_{l-1}\} + \{T_l - lT\} \\ &\qquad = \{(l-1)\Omega_1 + \cdots + 1\, \Omega_{l-1}\} + \text{const.} \end{aligned} \tag{5.14}$$

On neglecting the V-part and making use of (5.13) we obtain

$$T_l < lT + \frac{l(l-1)}{4}\cdot\kappa \log T \tag{5.15}$$

and hence

$$\log T_l < \log T + O(1).$$

This enables us to replace $\log T_l$ in (5.12) by $\log T$ and thus to prove by induction first

$$\Omega_l < \frac{\kappa}{2} \log T$$

and then (5.15). As to the V-part in (5.14), the equation with $l = k + 1$ yields the best harvest, namely

$$kV_1 + \cdots + 1V_k < (k+1)T + \frac{(k+1)k}{4}\cdot\kappa \log T,$$

a fortiori

$$V_l < \frac{k+1}{k-l+1}\, T + \frac{k(k+1)}{4(k-l+1)}\cdot\kappa \log T. \tag{5.16}$$

Roughly speaking, one may state these results (5.15), (5.16) as follows: It happens very rarely that $T_l(r)$ is appreciably bigger than $l\cdot T(r)$ or that $V_l(r)$ is appreciably bigger than $\dfrac{k(k+1)}{k-l+1}\, T(r)$ (the "rareness" referring to the values of r). *The mark for the transcendency level of the curve is set by the one order $T(r)$.*

6. Third Main Theorem

Our next goal is the relation

$$\Omega_1(r) + \sum_a m^*(r; a) = \tfrac{1}{2}q\kappa \log T(r) + \tfrac{1}{2}\omega(T). \tag{6.1}$$

The sum on the left extends over a set of q points $a^{(1)}, \cdots, a^{(q)}$ without accidental linear relations. Following in the footsteps of F. Nevanlinna and L. Ahlfors, we replace the surface element $d\tau_w$ on the w-sphere by $\mu_w \cdot d\tau_w$ with a certain density $\mu_w > 0$ in calculating the area $2\pi f(r; \alpha, \beta)$ of the Riemann surface over the w-sphere. μ may depend on α and β in addition to w. However, we shall assume from the outset that it is a homogeneous function only of the combinations ξ_i as introduced by (5.5):

$$\mu = \mu(\xi_0, \cdots, \xi_k).$$

The effect will be twofold: Q^2 in (5.4) will be replaced by $Q^2 \cdot \mu(\Xi_i)$ where Ξ_i arises from $\xi_i = \alpha_i w_2 - \beta_i w_1$ by substituting

$$\sum_j \alpha_j x_j(z), \qquad \sum_j \beta_j x_j(z)$$

for w_1, w_2 respectively; and the equation (5.8) will turn into

$$\Lambda(r) = \frac{1}{2}\mathop{\mathfrak{M}}_{\xi}\{N(r;\xi)\mu(\xi)\}. \tag{6.2}$$

We have

$$\Xi_i = \sum_j (\alpha_i\beta_j - \alpha_j\beta_i)x_j(z),$$

and (5.6) implies

$$\sum |\,\Xi_i\,|^2 = |\sum \alpha_i x_i\,|^2 + |\sum \beta_i x_i\,|^2.$$

On account of the inequality (3.5),

$$N(r;\xi) \leqq T(r) + m(r_0\,;\xi), \tag{6.3}$$

(6.2) will lead to

$$\Lambda(r) \leqq \tfrac{1}{2}\{CT(r) + C'\},$$

C, C' being the mean values

$$C = \mathop{\mathfrak{M}}_{\xi}\mu(\xi), \qquad C' = \mathop{\mathfrak{M}}_{\xi}\{\mu(\xi)m(r_0\,;\xi)\};$$

and instead of (5.9) we shall get

$$\frac{1}{2\pi}\int_0^{2\pi} \mathop{\mathfrak{M}}_{\alpha\beta} \log Q^2 \cdot d\varphi + \frac{1}{2\pi}\int_0^{2\pi} \mathop{\mathfrak{M}}_{\alpha\beta} \log \mu(\Xi_i)\cdot d\varphi = \omega(T).$$

We choose a continuous positive monotone function $g(r)$ defined for $0 < r \leqq 1$ which tends to infinity with $r \to 0$ in such a way that the integral

$$\int g(|\,z\,|)\,dz\,d\bar{z} = 2\pi\int_0^1 g(r)r\,dr$$

over the unit circle $|\,z\,| \leqq 1$ remains finite. An appropriate choice will be

$$g(r) = r^{-2}\left(\log\frac{c}{r}\right)^{-\varkappa} \tag{6.4}$$

with some constant $c > 1$. We then specify

$$\mu(\xi) = \prod_a g(||\,a\xi\,||). \tag{6.5}$$

On account of the singularity of $g(r)$ for $r = 0$ one has to examine carefully whether the integral

$$J = \int \mu(\xi)\cdot d\omega_\xi \tag{6.6}$$

over the ξ-sphere is *finite*. Once the finiteness of C is secured, C' will follow suit since $m(r_0\,;\,\xi)$ is a bounded function of the ξ. Taking the convergence of the integral J for granted, we obtain

$$2\Omega_1(r) + \sum_a \frac{1}{2\pi}\int_0^{2\pi} \mathop{\mathfrak{M}}_{\alpha,\beta} \log g(P_a)\cdot d\varphi = \omega(T)$$

with

$$P_a^2 = \frac{\left|\sum_{i<j}(\alpha_i\beta_j - \alpha_j\beta_i)(a_ix_j - a_jx_i)\right|^2}{\sum|a_i|^2\cdot(|\sum\alpha_ix_i|^2 + |\sum\beta_ix_i|^2)}.$$

In particular, for (6.4),

$$2\left\{\Omega_1(r) - \sum_a \frac{1}{2\pi}\int_0^{2\pi} \mathop{\mathfrak{M}}_{\alpha\beta} \log P_a\, d\varphi\right\}$$

(6.7)

$$= \omega(T) + \kappa\sum_a \frac{1}{2\pi}\int_0^{2\pi} \mathop{\mathfrak{M}}_{\alpha\beta} \log(\log c - \log P_a)\, d\varphi.$$

By the method often employed one gets

$$\mathop{\mathfrak{M}}_{\alpha\beta} \log P_a = \log[[ax]] - B$$

where B is the constant

$$B = -\mathop{\mathfrak{M}}_{\alpha\beta} \log\frac{|\alpha_0\beta_1 - \alpha_1\beta_0|}{\sqrt{|\alpha_0|^2 + |\beta_0|^2}} = B(k).$$

Hence the left side of (6.7), apart from an additive constant, equals twice the sum

$$\Omega_1(r) + \sum_a m^*(r; a).$$

For the individual term in the $\sum_a$ on the right side, we again resort to the concavity of the logarithm and we obtain

$$\frac{1}{2\pi}\int_0^{2\pi} \mathop{\mathfrak{M}}_{\alpha\beta} \log(\log c - \log P_a)\, d\varphi \leqq \log\frac{1}{2\pi}\int_0^{2\pi} \mathop{\mathfrak{M}}_{\alpha\beta}(\log c - \log P_a)\, d\varphi$$

$$= \log\{\text{const.} + m^*(r; a)\} \leqq \log\{\text{const.} + T(r)\}.$$

The resulting equation (6.1) holds good for the higher l. These *"defect" relations*

(6.8)

$$\boxed{\Omega_l(r) + \sum_a m_l^*(r; a) = \frac{1}{2}\omega(T_l) + \frac{\kappa q}{2}\log T_l}$$

constitute *the third* and most important *main theorem*. We repeat that *we assume the q points $a^{(1)}, \cdots, a^{(q)}$ to which the sum $\sum_a$ extends, to satisfy no acci-*

dental linear relation. Since $m_l^*(r;\ a)$ is a bounded function of r unless a is special, it is reasonable to restrict the formulas to the case where the a are $(l-1)$-spreads rather than complexes.

We still have to show the convergence of the integral (6.6). For that purpose we may without loss of generality assume $q \geqq k+1$. The q planes

$$(\xi a^{(1)}) = 0, \cdots, (\xi a^{(q)}) = 0 \tag{6.9}$$

cut the ξ-sphere up into a number of cells. According to our hypothesis of the linear independence of any $k+1$ of the forms (6.9), each vertex of the cell division is the intersection of exactly k of those planes, whose numbers $(p_1, \cdots, p_k)$ serve to characterize the vertex. We make use of the "dual" division where each vertex $(p_1, \cdots, p_k)$ is surrounded by a corresponding cell $\mathfrak{Z}(p_1, \cdots, p_k)$. For any point ξ of the ξ-sphere we arrange the q numbers $|(\xi a^{(p)})|$ in increasing order; ξ belongs to $\mathfrak{Z}(p_1, \cdots, p_k)$ if the k terms $p = p_1, \cdots, p_k$ (in any order) come first. We now determine an upper bound for the part $J(\mathfrak{Z})$ of the integral (6.6) extending over $\mathfrak{Z} = \mathfrak{Z}(1, \cdots, k)$. As each

$$\| \xi a^{(p)} \| \qquad (p = k+1, \cdots, q)$$

has a positive lower bound[2] in $\mathfrak{Z}$ we may at once cancel all the factors in (6.5) following the first k factors. By a linear transformation we make

$$(\xi a^{(1)}), \cdots, (\xi a^{(k)}); (\xi a^{(k+1)})$$

the coördinates $\xi_1, \cdots, \xi_k; \xi_0$. The Hermitian metric groundform will then no longer have the normal form $\sum |\xi_i|^2$. Of this change we take care by means of the constant γ/Γ mentioned in §1. The old argument $\| \xi a^{(1)} \|$ written in the new variables will then be

$$\geqq \frac{\gamma}{\Gamma} |\xi_1| : \{|\xi_0|^2 + \cdots + |\xi_k|^2\}^{-\frac{1}{2}}. \tag{6.10}$$

The non-homogeneous coördinates

$$\zeta_1 = \xi_1/\xi_0, \cdots, \zeta_k = \xi_k/\xi_0$$

are of modulus $\leqq 1$ within $\mathfrak{Z}$, and consequently (6.10) is

$$\geqq b\,|\zeta_1| \quad \text{with} \quad b = \frac{\gamma}{\Gamma} : \sqrt{k+1}$$

---

[2] In the most explicit way one sees this as follows. By introducing $(\xi a^{(1)}), \cdots, (\xi a^{(k)})$, $(\xi a^{(p)})$ as $k+1$ new independent variables and by expressing the ξ_i in terms of them one gets an inequality

$$\sum |\xi_i|^2 \leqq A\{|(\xi a^{(1)})|^2 + \cdots + |(\xi a^{(k)})|^2 + |(\xi a^{(p)})|^2\}.$$

Among the $k+1$ terms in the sum on the right side the last one is the biggest in $\mathfrak{Z}$; therefore

$$|(\xi a^{(p)})|^2 \geqq 1/(k+1)A \qquad \text{in } \mathfrak{Z}.$$

in $\mathfrak{Z}$. Hence we obtain for $J(\mathfrak{Z})$ an upper bound equal to a constant multiple of the k^{th} power of

$$\int_{|\zeta|\leq 1} g(b\,|\,\zeta\,|)\,d\zeta\,d\bar{\zeta} = \frac{2\pi}{b^2}\int_0^b g(r)r\,dr.$$

Appendix

We study a *degenerate* curve $\mathfrak{C}$ lying in a certain subspace R' of lower dimensionality h than R. After a suitable unitary transformation we may assume that this subspace is described by

$$x_{h+1} = 0, \cdots, x_k = 0.$$

The function elements $x_{h+1}(z), \cdots, x_k(z)$ vanish identically, whereas $x_0(z), \cdots, x_h(z)$ satisfy no linear relations with constant coefficients. For a given value r, $N(r;\alpha)$ will now depend in a homogeneous manner on the $h+1$ coefficients $\alpha_0, \cdots, \alpha_h$ only and will become singular, though staying bounded, for $\alpha_0 = \cdots = \alpha_h = 0$. The same will hold for $m(r;\alpha)$ *provided we define* $\|\alpha x\|^2$ *in* R', i.e. by

$$\left|\sum_{i=0}^{h}\alpha_i x_i\right|^2 : \sum_{i=0}^{h}|\alpha_i|^2 \cdot \sum_{i=0}^{h}|x_i|^2.$$

For points x in R' this differs from the corresponding expression in R only by the constant factor

$$\frac{|\alpha_0|^2 + \cdots + |\alpha_k|^2}{|\alpha_0|^2 + \cdots + |\alpha_h|^2}.$$

Hence the old $m(r;\alpha)$ will equal the new one plus half the logarithm of that constant. The modification has no influence upon $T(r)$. But it is important that the inequality (3.5) is correct with the *new* $m(r_0\,;\alpha)$ which is bounded with respect to the parameters α.

As a matter of fact we shall stick to the hypothesis that $\mathfrak{C}$ is non-degenerate or that R is the linear space of lowest dimension containing $\mathfrak{C}$. However, we must be prepared to have $\mathfrak{C}_l$ degenerate. Let R_l' be the sub-space of lowest dimensionality h_l containing $\mathfrak{C}_l$. We can *not* operate in R_l' throughout. In estimating Ω_l we had to use vectors α and β ranging over the k_l-sphere rather than an h_l-sphere; A_l in (5.11) is $= A(k_l)$, not $= A(h_l)$. But in the inequality (6.3),

$$N_l(r;\xi) \leqq T_l(r) + m_l(r_0\,;\xi),$$

and its subsequent application we make use of the *modified* m_l (defined in R_l') which is a bounded function of the parameters ξ. As to the final result (6.8) it is reasonable to limit oneself to points a in R_l' since $m_l^*(r;a) \sim 0$ for points outside R_l'.

The local investigation yielded as a byproduct a simple proof of the well-known fact: The Wronskian

$$\begin{vmatrix} x_0(z), & \cdots, x_k(z) \\ x_0'(z), & \cdots, x_k'(z) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ x_0^{(k-1)}(z), & \cdots, x_k^{(k-1)}(z) \end{vmatrix}$$

of $k+1$ analytic functions $x_0(z), \cdots, x_l(z)$ does not vanish identically unless they satisfy a non-trivial homogeneous linear relation with constant coefficients. Let $\alpha, \alpha', \cdots, \alpha^{(l-1)}$ be l planes in R. On applying the Wronskian theorem to the l functions $(\alpha x), \cdots, (\alpha^{(l-1)}x)$ one finds that $\mathfrak{C}_l$ *will not lie in any plane*

$$\sum \alpha_{i_1 \cdots i_l} x_{i_1 \cdots i_l} = 0$$

in R_l *whose coefficients* $\alpha_{i_1 \cdots i_l}$ *are special,*

$$\alpha_{i_1 \cdots i_l} = [\alpha \alpha' \cdots \alpha^{(l-1)}]_{i_1 \cdots i_l}.$$

From (2.7) it follows that $h_l + 1$ is at least as big as the number of numerically different l-sums

$$f_{i_1} + f_{i_2} + \cdots + f_{i_l} \qquad (i_1 < i_2 < \cdots < i_l) \tag{7.1}$$

which may be derived from $k+1$ given different numbers

$$f_0 < f_1 < \cdots < f_k.$$

By a fairly easy combinatorial argument[3] one thus concludes

$$h_l \geqq l(k - l + 1).$$

Simple examples show that the lower limit is actually attained for some curves $\mathfrak{C}$.

INSTITUTE FOR ADVANCED STUDY, PRINCETON, N. J.
UNIVERSITY OF ILLINOIS, URBANA, ILL.

---

[3] Indeed, let us string l beads on a vertical wire with (equidistant) halting points $0, 1, \cdots, k$ running from top to bottom ($k+1 \geqq l$). The beads, as they follow each other downwards, shall be labeled $1, \cdots, l$. They are movable and can occupy the nodes $i = 0, 1, \cdots, k$ on the wire. A position is described by the marks i_α which the several beads α occupy; $i_1 < i_2 < \cdots < i_l$. At the beginning of the game they shall be at the top: $i_1 = 0$, $i_2 = 1, \cdots, i_l = l - 1$. A simple move consists in moving one bead down one step while the others stay put. The problem is, by a series of simple moves to bring all the beads to the bottom where they will occupy the last l marks. It can be done, for instance, by first dropping the lowest bead, then the next, and so on. The number of simple moves needed is obviously $l(k - l + 1)$ because each bead has finally moved down $k - l + 1$ steps. If $(i_1, \cdots, i_l)$ are the $l(k - l + 1) + 1$ consecutive positions encountered in any such game, the corresponding number (7.1) increases by each move.

ANNALS OF MATHEMATICS
Vol. 39, No. 3, July 1938

THE ITERATED EXPONENTIAL INTEGERS

BY E. T. BELL

(Received January 17, 1938)

1. We define the iterated exponential integers $\xi_n^{(m)}$ $(m, n = 0, 1, \cdots)$ by

$$E_0(x) = \exp x,\ E_{m+1}(x) = \exp\,[E_m(x) - 1], \tag{1.1}$$

$$E_m(x) = \exp\,(\xi^{(m)} x) \equiv \sum_{n=0}^{\infty} \xi_n^{(m)} \frac{x^n}{n!}, \tag{1.2}$$

the exp ( ) in (1.2) being symbolic (after expansion, the n^{th} power of $\xi^{(m)}$ is replaced by $\xi_n^{(m)}$). Obviously, $E_n(x)$ is convergent for all finite x, and $\xi_0^{(m)} = \xi_n^{(0)} = 1$. In (1.10) it is seen that $\xi_n^{(m)}$ is an integer > 0. In §6 another system of intergers (not all positive), $\eta_n^{(m)}$, closely related to the $\xi_n^{(m)}$, is briefly noted. Most of the numbered statements in §§2–6 are theorems.

The $\xi_n^{(m)}$ have remarkable arithmetical properties, discussed in §§3–5. If these integers are arranged in an infinite rectangular array, $\xi_n^{(m)}$ being in row m, column n, the array has a species of double periodicity modulo p, p prime. The periodicity for n depends upon the expression of n in the scale of p, and is quite intricate; that for m depends upon n, but is otherwise simple. For given m, n the structure of the residue of $\xi_n^{(m)}$ mod p can be determined in a straightforward, non-tentative manner by the general theorems in §3; numerous applications of §3 are given in §§4, 5. Many of the special theorems are remarkably simple, for example, the striking result (4.5).

The $\xi_n^{(m)}$ are intimately connected with the Stirling numbers of both kinds: with those of the first kind through Lagrange's identical congruence (1.9); with those of the second kind through (1.10), proved in (2.7), which gives the structure of the array. Not enough is known of the arithmetical properties of the Stirling numbers to be of assistance with the $\xi_n^{(m)}$, and we obtain their arithmetic from the congruences in §3, which are characteristic for these numbers and the $\eta_n^{(m)}$. In investigating the arithmetic of the $\xi_n^{(m)}$, we find an apparently new algorithm, that of difference congruences, analogous to linear difference equations, the sufficient method.

With the exception of the $\xi_n^{(1)}$, the $\xi_n^{(m)}$ do not seem to have been considered before. The case $m = 1$ of $\xi_n^{(m)}$ suggests very little for $m > 1$, although it provides valuable checks on numerical calculations.

The $\xi_n^{(1)}$ are of interest in combinatorial analysis and elsewhere, and have been frequently investigated; their simpler properties have been rediscovered many times (1) (references at end of paper), by the authors cited, among several others. Incidentally, the numbers K_r^m of d'Ocagne (1) are merely Stirling

numbers (2) of the second kind, and all of the results in his paper had been obtained by numerous English writers (following Herschel) about thirty years earlier, together with many further properties. Dobiński's result of 1877,

$$\sum_{s=1}^{\infty} \frac{s^n}{s!} = e\xi_n^{(1)},$$

where e has its usual meaning, has been ascribed to Euler, but without a specific reference. It is an immediate consequence of (1.10), and there are similar (but more complicated) summations involving $\xi_n^{(m)}$, $m > 1$. Catalan's problem of 1838 (loc. cit. (1), p. 508–16, vol. 3) of finding the number of different products of n factors is similar to Epstein's (1) of finding the total number of decompositions of an integer into coprime factors where, instead of Catalan's sequence 1, 2, 5, 14, 42, 132, $\cdots$, the sequence where 1, 2, 3, $\cdots$ distinct primes are concerned is 1, 2, 5, 15, 52, 203, $\cdots$, that is, $\xi_n^{(1)}$. Catalan's numbers first appeared in Euler's and Segner's problem of finding the number of partitions of a rectilinear n-gon into triangles by diagonals. In my previous papers (1) the connection between the $\xi_n^{(1)}$ and partitions was noted. A combinatorial interpretation of $\xi_n^{(1)}$ is given by Whitworth (loc. cit. (1), p. 97); a readily grasped interpretation of the same sort is that $\xi_n^{(1)}$ is the total number of possible rhyme-schemes for a stanza of n verses. Some of Touchard's (1) results for $\xi_n^{(1)}$ are noted presently.

The definition and notation in (1.3)–(1.8), (1.16) *are fixed in the sequel.*

(1.3) m, n, r, s, p are non-negative integers; p *is prime*; x, y are real or complex variables.

(1.4) Δ is the operator of finite differences;

$$\Delta\phi(x) \equiv \phi(x+1) - \phi(x);$$

$$\Delta^m 0^n \equiv (\Delta^m x^n)_{x=0},\ mn \neq 0.$$

(1.5) A polynomial in x, with rational coefficients, which takes only integer values > 0 for all integer values $\geqq 0$ of x, will be called *integral.*

The $\Delta^m 0^n/m!$, $mn \neq 0$, are the Stirling numbers of the second kind; they are integers > 0 if $m \leqq n$, and are equal to zero if $m > n$. The Stirling numbers of the first kind are the coefficients in the expansion of

$$(1.6) \qquad (x)_n \equiv x(x-1)\cdots(x-n+1),\ n > 0.$$

To avoid trivial exceptions we define $(x)_0 \equiv 1$. Note that Jordan's (2) notation $(x)_n$ for the usual factorial $x^{(n)}$ is used here to prevent confusion with $\xi^{(m)}$. Thus (an important instance for the sequel),

$$(1.7) \qquad (\xi^{(m)})_n \equiv \xi^{(m)}(\xi^{(m)} - 1)\cdots(\xi^{(m)} - n + 1), \qquad n > 0,$$

in which the r^{th} power of $\xi^{(m)}$ is to be replaced by $\xi_r^{(m)}$ after expansion of the right, according to the usual procedure in the symbolic (or umbral) calculus (3); $(\xi^{(m)})_0 \equiv \xi_0^{(m)}, = 1$.

(1.8) $f(x)$ denotes an arbitrary polynomial in x, or, if the series is convergent, a power series in x.

In using the umbral calculus, all indicated algebraic operations must be completed before exponents are degraded to suffixes, and the exponents 0, 1 must be treated precisely as exponents > 1. For example, $\xi^{(m)}$ does not denote $(\xi^{(m)})^1$, which is $\xi_1^{(m)}$, but is the umbra of the sequence $\xi_n^{(m)}$ $(n = 0, 1, \cdots)$, or of the one-rowed matrix $(\xi_0^{(m)}, \xi_1^{(m)}, \cdots)$. Similarly, $\xi_0^{(m)}$ is written $(\xi^{(m)})^0$, and is 1, but not for the algebraic reason that $x^0 = 1$.

Extensive use will be made of Lagrange's identical congruence

$$(x)_p \equiv x^p - x \bmod p, \tag{1.9}$$

which will connect the $\xi_n^{(m)}$ with the Stirling numbers of the first kind; the connection with those of the second kind is given by

$$\xi_n^{(m+1)} = \sum_{s=1}^{n} \xi_s^{(m)} \frac{\Delta^s 0^n}{s!}, \qquad n > 0, \tag{1.10}$$

proved and generalized in (2.7). As already remarked, (1.10), with $\xi_0^{(m)} = \xi_n^{(0)} = 1$, gives the structure of the array for $\xi_n^{(m)}$. It is to be noted that the proofs of the theorems on congruences go through only because (1.9) is *identical* ($\equiv$), that is, coefficients of like powers of x on opposite sides of $\equiv$ are respectively congruent modulo p. The congruences for the $\xi_n^{(m)}$, derived from extensions of (1.9), are not identical; hence the notation $\equiv \bmod p$ is used for such congruences. Touchard's (1) congruences for the $\xi_n^{(1)}$, in particular

$$\xi_{p+n}^{(1)} \equiv \xi_n^{(1)} + \xi_{n+1}^{(1)} \bmod p, \tag{1.101}$$

from which all for the $\xi_n^{(1)}$ follow, are obtained incidentally as special cases ($m = 1$) of the general congruences for the $\xi_n^{(m)}$, and have been used to furnish numerical checks. It will be seen however in (2.6), (3.4), that (1.101) is so special as to be anomalous in the general theory.

We give here for future use some simple consequences of (1.9), and prove first that

$$(x)_{mp+n} \equiv (x^p - x)^m (x)_n \bmod p. \tag{1.11}$$

The congruence holds for $m = 0$. If $m > 0$, without loss of generality we may assume $0 \leqq n < p$. Suppose first that $n > 0$. Then

$$\begin{aligned}
(x)_{mp+n} &= \prod_{t=0}^{mp+n-1} (x - t), \\
&= \prod_{t=0}^{m-1} [(x - tp) \cdots (x - tp - p + 1)] \prod_{u=0}^{n-1} (x - mp - u), \\
&\equiv [x(x-1) \cdots (x - p + 1)]^m \prod_{u=0}^{n-1} (x - u) \bmod p, \\
&\equiv [(x)_p]^m (x)_n \bmod p, \\
&\equiv (x^p - x)^m (x)_n \bmod p,
\end{aligned}$$

the last by (1.9). If $n = 0$, the second product is absent, and (1.11) holds in this case also; which completes the proof.

From (1.11) we get the reduced form of the most general case. By the congruence properties of binomial coefficients,

$$\binom{p}{r} \equiv 0 \bmod p, \qquad 0 < r < p; \qquad \binom{p}{0} \equiv \binom{p}{p} \equiv 1 \bmod p;$$

whence, by complete induction,

$$(x^p - x)^{p^m} \equiv x^{p^{m+1}} - x^{p^m} \bmod p. \tag{1.12}$$

Applying this to (1.11) we find

(1.13) *If* $m = \sum_{j=0}^{t} m_j p^{t-j}$ *is the expression of* m *in the scale of* p,

$$(x)_m \equiv (x)_{m_t} \prod_{j=0}^{t-1} (x^{p^{t-j}} - x^{p^{t-j-1}})^{m_j} \bmod p,$$

the product being vacuous if $t = 0$.

If the coefficients $f(0)$, $\Delta^s f(0)/s!$, $s > 0$, in the factorial expansion

$$f(x) = f(0) + \sum_{s=1} \frac{\Delta^s f(0)}{s!} (x)_s$$

of $f(x)$ are integers, (1.11) gives

$$\frac{\Delta^s f(0)}{s!} (x)_{rp+s} \equiv \frac{\Delta^s f(0)}{s!} (x^p - x)^r (x)_s \bmod p,$$

and therefore

$$\begin{aligned} \sum_{s=1} \frac{\Delta^s f(0)}{s!} (x)_{rp+s} &\equiv (x^p - x)^r \sum_{s=1} \frac{\Delta^s f(0)}{s!} (x)_s \bmod p, \\ &\equiv (x^p - x)^r [f(x) - f(0)] \bmod p. \end{aligned}$$

Using (1.11) again we get

(1.14) *If* $f(0)$, $\Delta^s f(0)/s!$, $s > 0$, *are integers,*

$$(x^p - x)^r f(x) \equiv (x)_{rp} f(0) + \sum_{s=1} \frac{\Delta^s f(0)}{s!} (x)_{rp+s} \bmod p.$$

Consider now an integral polynomial $P(x)$, as defined in (1.5). We may write $P(x)$ in the canonical form $(a/b)x^c\, Q(x)$, where a, b are coprime integers > 0, c is an integer $\geqq 0$, and $Q(x)$ is a polynomial in x, not algebraically divisible by x, with integer coefficients whose G. C. D. is 1, which takes integer values > 0 for integer values $\geqq 0$ of x, by (1.5). Hence

$$\begin{aligned} P(x + y) &= \frac{a}{b} (x + y)^c\, Q(x + y), \\ &= \frac{a}{b} [x^c + yS(x, y)] [Q(x) + yT(x, y)], \end{aligned}$$

in which $S(x, y)$, $T(x, y)$ are polynomials in x, y with integer coefficients. Thus

$$P(x + y) = P(x) + y\left[\frac{a}{b} R(x, y)\right]$$

the term in [ ] being a polynomial in x, y with rational coefficients which takes integer values for integer values $\geqq 0$ of x, y since $P(x + y)$, $P(x)$ take integer values > 0 for integer values $\geqq 0$ of x, y; the coefficients of $R(x)$ are integers. Hence we have

(1.15) *If the integral polynomial $P(x)$ is $(a/b)x^c\, Q(x)$ when written in canonical form as above, and if p (prime, as always) is not a divisor of b, then*

$$P(rp + s) \equiv P(s) \bmod p.$$

(1.16) A prime which divides b in (1.15) is called *exceptional* for $P(x)$.

2. In this section we obtain the fundamental recurrences from which the $\xi_n^{(m)}$ may be calculated and from which, in conjunction with (1.13), (1.14), arithmetical properties will be derived in §§3–5. By (1.1), (1.2),

$$\exp\,(\xi^{(m+1)}x) = \exp\,[\exp\,(\xi^{(m)}x) - 1];$$

whence, by differentiation with respect to x,

$$\xi^{(m+1)} \exp\,(\xi^{(m+1)}x) = \xi^{(m)} \exp\,[(\xi^{(m+1)} + \xi^{(m)})x].$$

Equating coefficients of x^n we have

$$\xi_{n+1}^{(m+1)} = \xi^{(m)}(\xi^{(m+1)} + \xi^{(m)})^n, \tag{2.1}$$

and thence, by the usual processes of the umbral calculus,

$$\xi^{(m+1)} f(x + \xi^{(m+1)}) = \xi^{(m)} f(x + \xi^{(m)} + \xi^{(m+1)}), \tag{2.2}$$

of which (2.1) is a special case.

Applying Herschel's expansion

$$f(e^x) = f(1) + \sum_{n=1}^{\infty} \frac{x^n}{n!} f(1 + \Delta)0^n$$

to (1.1), (1.2), we have

$$\exp\,(\xi^{(m+1)}x) = 1 + \sum_{n=1}^{\infty} \frac{x^n}{n!} (\exp\,(\xi^{(m)}\Delta))0^n.$$

Equating coefficients of x^n in the last gives (1.10).

The next is obtained by the symbolic method in finite differences and the umbral calculus, but may be proved at greater length otherwise. If $|\, x\,| < 1$, we may replace x by $\log\,(1 + x)$ in (1.1). Then

$$(1 + x)^{\xi^{(m+1)}} = \exp\,(\xi^{(m)}x),$$

and hence, by equating coefficients of x^n

$$(\xi^{(m+1)})_n = \xi_n^{(m)}, \tag{2.3}$$

the notation on the left being the factorial in (1.7) with m replaced by $m+1$. This relation is fundamental for the arithmetical properties of the $\xi_n^{(m)}$.

A radical distinction between the $\xi_n^{(1)}$ and the $\xi_n^{(m)}$, $m > 1$, may be noted here. This is rooted in the fact that $\xi_n^{(0)} = 1$, and hence $\xi_s^{(0)}\xi_n^{(0)} = \xi_{s+n}^{(0)}$. But $\xi_s^{(m)}\xi_n^{(m)} \neq \xi_{s+n}^{(m)}$ if $m > 0$. By the previous remark, (2.2), (2.3) for $m = 0$ become

$$\xi^{(1)}f(x + \xi^{(1)}) = f(x + 1 + \xi^{(1)}), \qquad (\xi^{(1)})_n = 1. \tag{2.4}$$

For $x = 0$, the first of (2.4) gives

$$f(\xi^{(1)} + 1) = \xi^{(1)}f(\xi^{(1)}). \tag{2.5}$$

Write $\xi^{(1)} \equiv \alpha$ for the moment, and consider the function $F(\alpha) \equiv (\alpha)_n f(\alpha)$. Then (2.5) is $\alpha f(\alpha) = f(\alpha + 1)$. Applying this to

$$F(\alpha) \equiv \alpha(\alpha - 1) \cdots (\alpha - n + 1)f(\alpha),$$

we have

$$F(\alpha) = \alpha(\alpha - 1) \cdots (\alpha - n + 2)f(\alpha + 1),$$

and hence, by repeated applications in the same way, finally $F(\alpha) = f(\alpha + n)$. That is,

$$(\xi^{(1)})_n f(\xi^{(1)}) = f(\xi^{(1)} + n), \tag{2.6}$$

a functional equation for $\xi^{(1)}$. There is no similar functional equation for $\xi^{(m)}$, $m > 1$. In (2.6) take $n = p$ and apply (1.9):

$$[(\xi^{(1)})^p - \xi^{(1)}]f(\xi^{(1)}) \equiv f(\xi^{(1)} + p) \bmod p,$$

if f is a polynomial with integer coefficients. Choosing $f(x) \equiv x^n$ in this, we get (1.101).

Continuing with the general case: (2.3) and the expansion of $f(x)$ in factorials give

$$f(\xi^{(m+1)}) = f(0) + \sum_{s=1} \frac{\Delta^s f(0)}{s!}\, \xi_s^{(m)}, \tag{2.7}$$

which becomes (1.10) for $f(x) \equiv x^n$.

The following give the algebraic structure of the $\xi_n^{(m)}$.

$$\xi_n^{(0)} = \xi_0^{(m)} = \xi_1^{(m)} = 1; \tag{2.8}$$

(2.9) *$\xi_n^{(m)}$, $n > 1$, is an integral polynomial of degree $n - 1$ in m, and is algebraically divisible by $m + 1$:*

$$\xi_n^{(m)} = c(m + 1)\, P_n(m), \qquad n > 1,$$

in which c is a rational number, if fractional in its lowest terms, and $P_n(m)$ is a polynomial of degree $n - 2$ in m with integer coefficients.

In connection with the congruences we shall use the following instances of (2.9), proved presently:

$$\xi_2^{(m)} = m + 1,$$

$$\xi_3^{(m)} = \tfrac{1}{2}(m + 1)(3m + 2),$$

$$\xi_4^{(m)} = \tfrac{1}{2}(m + 1)(2m + 1)(3m + 2)$$

$$\xi_5^{(m)} = \tfrac{1}{6}(m + 1)(45m^3 + 70m^2 + 35m + 6),$$

$$\xi_6^{(m)} = \tfrac{1}{24}(m + 1)(540m^4 + 1005m^3 + 665m^2 + 202m + 24).$$

The assertions in (2.9) are proved by mathematical induction from (2.1), (2.8). The first two of (2.8) are already known from §1. The third follows on taking $n = 0$ in (2.1), when $\xi_1^{(m+1)} = \xi_1^{(m)}$, a difference equation whose solution is $\xi_1^{(m)} = 1$, since $\xi_n^{(0)} = 1$, and hence in particular, $\xi_1^{(0)} = 1$. Thus we attend only to (2.1) with $m > 0$, $n > 1$. The expanded form may be written in this case

$$\xi_n^{(m)} - \xi_n^{(m-1)} = \sum_{s=1}^{n-1} \binom{n-1}{s} \xi_{n-s}^{(m-1)} \xi_s^{(m)}, \qquad m > 0, \qquad n > 1, \tag{2.11}$$

with the usual notation for binomial coefficients. In (2.11) we take $n = 2, 3, \cdots$, and solve the resulting difference equations, at each step using the results of previous solutions and (2.8). Thus for $n = 2$,

$$\xi_2^{(m)} - \xi_2^{(m-1)} = \xi_1^{(m-1)} \xi_1^{(m)} = 1;$$

replacing m by $1, 2, \cdots, m$ in this, and adding the results, we get

$$\xi_2^{(m)} - \xi_2^{(0)} = m,$$

and therefore, by (2.8), the first of (2.10).

For $n = 3$, (2.11) and the first of (2.10) give

$$\xi_3^{(m)} - \xi_3^{(m-1)} = 3m + 1,$$

whose solution is found as before, and is the second of (2.10). In the same way,

$$\xi_4^{(m)} - \xi_4^{(m-1)} = 9m^2 + 4m + 1,$$

and, generally, for $n > 1$,

$$\xi_n^{(m)} - \xi_n^{(m-1)} = Q_n(m) \tag{2.12}$$

where $Q(m)$ is an integral polynomial of degree $n - 2$ in m; the mathematical induction to prove (2.12) is immediate from (2.11) and the first of (2.10). Since the solution $\xi_n^{(m)}$ of (2.12) is a polynomial in m of degree $n - 1$, by the elementary properties of sums $1^n + 2^n + \cdots + m^n$, and since $\xi_n^{(m)}$ is an integer, all of (2.9) except the algebraic divisibility by $m + 1$ is proved. From (2.12) and the first of (2.10) the divisibility follows by mathematical induction.

It follows from (2.8), (2.9) that the n^{th} term $\xi_n^{(m)}$ of every sequence $\xi^{(m)}$,

$m \geqq 0$, is known when the n^{th} terms $\xi_n^{(s)}$, $s = 0, \cdots, n-1$, of the first n sequences are known. For, by (2.9),

$$\xi_n^{(m)} = c_{n-1}m^{n-1} + c_{n-2}m^{n-2} + \cdots + c_0, \tag{2.13}$$

in which $c_{n-1}, \cdots, c_0$ are constant rational numbers. In this take $m = 0, 1, \cdots, n-1$. Then $c_0 = \xi_n^{(0)} = 1$; the equations

$$m^{n-1}c_{n-1} + m^{n-2}c_{n-2} + \cdots + mc_1 = \xi_n^{(m)} - 1, \; m = 1, \cdots, n-1 \tag{2.14}$$

have non-vanishing determinant, and hence are solvable for $c_{n-1}, \cdots, c_1$. The solution sets an upper bound to the denominator of c in (2.9).

From (2.13) we see

(2.15) *For n fixed, > 1, $\xi_n^{(m)}$ satisfies a linear difference equation of order $n-1$ with constant rational coefficients,*

$$a_1\xi_n^{(m+n-1)} + a_2\xi_n^{(m+n-2)} + \cdots + a_n\xi_n^{(m)} = a_{n+1}.$$

Thus, for $n = 3$ the equation is

$$\xi_3^{(m+2)} - 2\xi_3^{(m+1)} + \xi_3^{(m)} = 3,$$

obtained from (2.13) with $n = 3$, on replacing m by m, $m+1$, $m+2$, and eliminating m^2, m from the resulting equations (linear in m^2, m). Similarly for any $n > 1$. The elimination can also be performed by using the recurrence relations for sums of like powers of the natural numbers.

3. The fundamental congruences for the $\xi_n^{(m)}$ are (3.1), (3.7), (3.12). In (1.14) replace x by $\xi^{(m+1)}$ and apply (2.3) to the right of the result. Then

(3.1) *If $f(0)$, $\Delta^s f(0)/s!$, $s > 0$, are integers,*

$$[(\xi^{(m+1)})^p - \xi^{(m+1)}]^r f(\xi^{(m+1)}) \equiv \xi_{rp}^{(m)} f(0) + \sum_{s=1} \frac{\Delta^s f(0)}{s!}\, \xi_{rp+s}^{(m)} \bmod p.$$

All indicated algebraic operations on the left are to be performed before exponents are degraded. As in the proof of (1.13), [ ] on the left can be reduced immediately, but we need not write out the result. For $f(x) \equiv x^n$, $n > 0$, there is the useful special case

$$[(\xi^{(m+1)})^p - \xi^{(m+1)}]^r (\xi^{(m+1)})^n \equiv \sum_{s=1}^{n} \frac{\Delta^s 0^n}{s!}\, \xi_{rp+s}^{(m)} \bmod p. \tag{3.2}$$

For $r = 0$ this gives

$$\xi_n^{(m+1)} \equiv \sum_{s=1}^{n} \frac{\Delta^s 0^n}{s!}\, \xi_s^{(m)} \bmod p, \qquad n > 0, \tag{3.3}$$

which is obvious from (1.10); for $r = 1$,

$$\xi_{p+n}^{(m+1)} - \xi_{n+1}^{(m+1)} \equiv \sum_{s=1}^{n} \frac{\Delta^s 0^n}{s!}\, \xi_{p+s}^{(m)} \bmod p, \qquad n > 0. \tag{3.4}$$

To complete (3.4), the following immediate consequence of (1.9), (2.3) is required

$$\xi_p^{(m+1)} - \xi_1^{(m+1)} \equiv \xi_p^{(m)} \mod p, \tag{3.5}$$

a difference congruence for the residue of $\xi_p^{(m)}$ modulo p, whose solution is contained in the more general (4.4).

In (3.1) take $f(x) \equiv (x)_n$, $n > 0$, and recall that $\Delta^s(x)_n = (n)_s(x)_{n-s}$, $ns \neq 0$. Then all coefficients on the right vanish, except the last ($=1$), and we have

$$[(\xi^{(m+1)})^p - (\xi^{(m+1)})]^r(\xi^{(m+1)})_n \equiv \xi_{rp+n}^{(m)} \mod p, \tag{3.6}$$

which also follows at once from (1.11), (2.3). If $n = 0$, the symbolic factorial on the left may be dropped, by (1.11). From the special case $r = 1$ of (3.6), (3.1) can be deduced. For $r = 1$, $n = 0$, (3.6) gives (3.5).

Here again we may note the exceptional character of the $\xi_n^{(1)}$. From (3.4), (3.5), (1.10) for $m = 0$, and $\xi_n^{(0)} = 1$, we get (1.101); but there is no similar reduction of (3.4) if $m > 0$.

From (2.1) we obtain another type of congruences for the $\xi_n^{(m)}$, by proceeding from (2.1) as in deriving (1.13):

$$\xi_{rp+s+1}^{(m+1)} \equiv \xi^{(m)}[(\xi^{(m+1)})^p + (\xi^{(m)})^p]^r[\xi^{(m+1)} + \xi^{(m)}]^s \mod p. \tag{3.7}$$

To prevent possible slips in working with (3.7), we repeat the caution that all indicated algebraic operations must be completed before exponents are degraded. For example, from (3.7) with r replaced by p^r, $s = 0$, and (2.8), we get the interesting result

$$\xi_{p^{r+1}+1}^{(m+1)} - \xi_{p^{r+1}+1}^{(m)} \equiv \xi_{p^{r+1}}^{(m+1)} \mod p. \tag{3.8}$$

From (3.6) for $n = 1$ and r replaced by p^r, we find

$$\xi_{p^{r+1}+1}^{(m+1)} - \xi_{p^{r+1}+1}^{(m)} \equiv \xi_{p^{r+1}}^{(m+1)} \mod p,$$

and hence, by (3.8),

$$\xi_{p^{r+1}}^{(m)} \equiv \xi_{p^{r+1}}^{(m)} \mod p, \tag{3.9}$$

$m + 1$ being replaced by m finally, since (3.9) holds for $m = 0$. This example is completed in (4.5), (4.6).

To complete (3.2) we need the following immediate consequence of (1.11), (2.3):

$$[(\xi^{(m+1)})^p - (\xi^{(m+1)})]^r \equiv \xi_{rp}^{(m)} \mod p. \tag{3.10}$$

In this replace r by p^r. Then

$$\xi_{p^{r+1}}^{(m+1)} - \xi_{p^r}^{(m+1)} \equiv \xi_{p^{r+1}}^{(m)} \mod p. \tag{3.11}$$

If after expansion by MacLaurin's theorem of the functions in (2.2), x be replaced by p, we get

(3.12) *If $P(x)$ is an integral polynomial in which no denominator of a non-integral coefficient (if any) is divisible by p,*

$$\xi^{(m+1)}P(\xi^{(m+1)}) \equiv \xi^{(m)}P(\xi^{(m)} + \xi^{(m+1)}) \mod p.$$

4. The general congruences in §3 are sufficient for a detailed investigation of the residue mod p of $\xi_n^{(m)}$. In this section and the following we shall obtain several specific results, from which the general procedure will be evident. The congruences in this section refer to the suffix n; the superscript m is considered in §5. To avoid repetition, (2.8) will be used where necessary without further reference, and all polynomials occuring on the right of congruences are integral (by the method of their derivation). If in working with one of the general congruences in §3, m is replaced by $m - 1$ for convenience, the new m is > 0 until the final statement, when it is seen by inspection that the result holds also for $m = 0$, by (2.8). The notation p, m, n, r, s is as always; see (1.3).

We prove first

$$\xi_{(r+1)p+s}^{(m)} \equiv P_{r,s}(m) \mod p, \qquad 0 \leqq s < p, \tag{4.1}$$

where $P_{r,s}(m)$ is an integral polynomial in m of degree $2r + s + 1$;

(4.2) $\xi_{(r+1)p} \equiv Q_r(m) \mod p$, *where $Q_r(m)$ is an integral polynomial in m of degree $2r + 1$.*

The proof will be by double induction on r, s from (2.9), (3.2), (3.6), (3.7), (3.10) and the following special cases, proved first, all of which will be used later.

$$\begin{aligned}
Q_0(m) &\equiv m + 1,\\
Q_1(m) &\equiv \tfrac{1}{6}(m + 1)(m + 2)(2m + 3),\\
Q_2(m) &\equiv \tfrac{1}{40}(m + 1)(m + 2)(8m^3 + 31m^2 + 41m + 20);\\
P_{0,1}(m) &\equiv \tfrac{1}{2}(m + 1)(m + 2),\\
P_{0,2}(m) &\equiv \tfrac{1}{6}(m + 1)(m + 2)(4m + 3),\\
P_{1,1}(m) &\equiv \tfrac{1}{6}(m + 1)^2(m + 2)(2m + 3),\\
P_{0,3}(m) &\equiv \tfrac{1}{12}(m + 1)(m + 2)(3m + 2)(5m + 3),\\
P_{1,2}(m) &\equiv \tfrac{1}{120}(m + 1)(m + 2)(68m^3 + 211m^2 + 201m + 60).\\
P_{2,1}(m) &\equiv \tfrac{1}{120}(m + 1)(m + 2)(34m^4 + 159m^3 + 275m^2 + 212m + 60).
\end{aligned} \tag{4.3, 4.4}$$

In all proofs repeated use is made of the fact that $1^m + 2^m + \cdots + n^m$, $nm \neq 0$, when expressed as a Bernoulli polynomial, is an integral polynomial in n of degree $m + 1$. This need not be mentioned again, particularly in the inductive proofs. These Bernoulli sums for $m = 1, \cdots, 10$ are available in many places, and will be used in solving difference congruences without further reference.

In (3.11) replace r by $0, 1, \cdots, r$, add, and in the result replace r by $r - 1$. Then

$$\xi_{p^r}^{(m+1)} \equiv \sum_{s=0}^{r} \xi_{p^s}^{(m)} \mod p.$$

This may be solved precisely as a difference equation, m being the variable. Thus, for $m = 0$ we get

$$\xi_{p^r}^{(1)} \equiv r + 1 \mod p.$$

Taking $m = 1$, and using the last, we get the congruence for $\xi_{p^r}^{(2)}$, and so on; finally

$$\xi_{p^r}^{(m)} \equiv \binom{m+r}{r} \mod p. \tag{4.5}$$

For $r = 1$, this is the first of (4.3), and (4.5) is a generalization of (4.2) for $r = 0$. From (3.9), (4.5) we get

$$\xi_{p^r+1}^{(m)} \equiv \binom{m+r+1}{r+1} \mod p, \tag{4.6}$$

which for $r = 1$ is the first of (4.4), and which generalizes (4.1) for $r = 0$, $s = 1$; (4.6) is generalized in (4.24).

In (3.6) take $n = 0$, $r = 2$, in order to get a difference congruence for $\xi_{2p}^{(m)}$. Then

$$\xi_{2p}^{(m+1)} - 2\xi_{p+1}^{(m+1)} + \xi_2^{(m+1)} \equiv \xi_{2p}^{(m)} \mod p.$$

The value of $\xi_2^{(m+1)}$ is obtained from (2.10) and is $m + 2$; that of $\xi_{p+1}^{(m+1)} \pmod p$, from (4.6). Hence

$$\xi_{2p}^{(m+1)} - \xi_{2p}^{(m)} \equiv (m+2)^2 \mod p,$$

which is solved by replacing m by $0, 1, \cdots, m-1$, adding the results, and noting that the congruence thus obtained holds also for $m = 0$. This gives the second of (4.3). A difference congruence for $\xi_{p+2}^{(m)}$ is obtained from (3.2) with $r = 1$, $n = 2$:

$$\xi_{p+2}^{(m+1)} - \xi_{p+2}^{(m)} \equiv \xi_{p+1}^{(m)} + \xi_3^{(m+1)} \mod p,$$

and hence by (2.10) and the first of (4.3),

$$\xi_{p+2}^{(m+1)} - \xi_{p+2}^{(m)} \equiv (m+2)(2m+3) \mod p.$$

Solved as in the preceding example, this gives the second of (4.4).

It will be helpful for the final induction to inspect an alternative derivation from (3.7) of the last. Taking $r = s = 1$ in (3.7), we get

$$\xi_{p+2}^{(m+1)} - \xi_{p+2}^{(m)} \equiv \xi_{p+1}^{(m)} + \xi_{p+1}^{(m+1)} + \xi_2^{(m)}\xi_p^{(m+1)} \mod p,$$

the right of which, by the first of (4.3), (4.4), already proved, and (2.10), is congruent mod p to an integral polynomial in m of degree 2. Hence the solution of the difference congruence will be an integral polynomial of degree 1 higher, namely 3. Similarly in the general case. The right of the foregoing congruence reduces to $(m+2)(2m+3) \bmod p$, hence providing a check.

For the third of (4.4) we take $r = 2$, $s = 0$ in (3.7) with m replaced by $m - 1$:

$$\xi_{2p+1}^{(m)} - \xi_{2p+1}^{(m-1)} \equiv \xi_{2p}^{(m)} + 2\xi_{p+1}^{(m-1)}\xi_p^{(m)} \mod p,$$

and hence, by previous results,

$$\xi_{2p+1}^{(m)} - \xi_{2p+1}^{(m-1)} \equiv \tfrac{1}{6}(m+1)(8m^2 + 13m + 6) \mod p,$$

of which the solution, found as before, is as stated in (4.4).

The third of (4.3) is now obtained from (2.10) with $r = 3$ and m replaced by $m - 1$:

$$\xi_{3p}^{(m)} - \xi_{3p}^{(m-1)} \equiv 3\xi_{2p+1}^{(m)} - 2(m+1)^3 \mod p,$$

on reducing by previous results; hence, by the third of (4.4),

$$\xi_{3p}^{(m)} - \xi_{3p}^{(m-1)} \equiv \tfrac{1}{2}(m+1)^2(2m^2 + 3m + 2) \mod p,$$

solved as before to give the third of (4.3).

The fourth of (4.4) comes from (3.7) with $r = 2$, $s = 1$ in a similar manner; the step next to the last is

$$\xi_{2p+2}^{(m)} \equiv \tfrac{1}{6}\sum_{s=0}^{m}(s+1)(17s^3 + 32s^2 + 23s + 6) \mod p.$$

The proof by mathematical induction as outlined after (4.2) is immediate on noting the degrees of the integral polynomials on the right of the difference congruences to be solved on the assumption that (4.1), (4.2) hold up to any prescribed r, s. Thus from (3.10) we have

$$\xi_{tp}^{(m)} - \xi_{tp}^{(m-1)} \equiv \sum_{s=1}^{t}(-1)^s\binom{t}{s}\xi_{(t-s)p+s}^{(m)} \mod p;$$

and hence if (4.1) holds with $r = 0, 1, \cdots, t-2$, $t \geqq 2$,

$$\xi_{tp}^{(m)} - \xi_{tp}^{(m-1)} \equiv -(-1)^t\xi_t^{(m)} - \sum_{s=1}^{t-1}(-1)^s\binom{t}{s}P_{t-s-1,s}^{(m)} \mod p,$$

the right of which is an integral polynomial in m of degree $2t - 2$. The solution of the last is therefore $\xi_t^{(m)} \equiv R(m) \mod p$, where $R(m)$ is an integral polynomial in m of degree $2t - 1$. Hence (4.2) is proved on the assumption that (4.4) holds for the values of r specified above. From the expanded form of (3.7) corresponding to the special cases in proofs of (4.4), we prove (4.1) by a straightforward double induction on r, s, keeping r fixed at a particular step. From the inductions we have also

(4.7) *Each of* $P_{r,s}(m)$, $Q_r(m)$ *in* (4.1), (4.2) *is algebraically divisible by* $m + 1$.

The most general result of the type considered in this section refers to the residue mod p of $\xi_n^{(m)}$, $n > 0$, when n is expressed as in (1.13) in the scale of p. By (2.9), a crude form of the residue is given at once as a polynomial of degree $n - 1$, and it is theoretically possible, but impracticable, to reduce this crude degree by repeated use of Fermat's theorem and the arithmetical properties of

binomial coefficients. The crude degree is much too high; in (4.6), for example, it is $p^r + 1$, while the reduced degree is $r + 1$. In obtaining further results, it is convenient to take $m > 0$ till the final step, as explained before, and write $\xi^{(m)} \equiv \beta$, $\xi^{(m-1)} \equiv \alpha$ in (3.2), (3.6), (3.7), (3.10), (4.5), (4.6), which thus become

$$(\beta^p - \beta)^r \beta^s \equiv \sum_{u=1}^{s} \frac{\Delta^u 0^s}{u!} \alpha_{rp+u} \mod p, \tag{4.8}$$

$$(\beta^p - \beta)^r (\beta)_s \equiv \alpha_{rp+s} \mod p, \tag{4.9}$$

$$\beta_{rp+s+1} \equiv \alpha(\alpha + \beta)^s (\alpha^p + \beta^p)^r \mod p, \tag{4.10}$$

$$(\beta^p - \beta)^r \equiv \alpha_{rp} \mod p, \tag{4.11}$$

$$\beta_{p^r} \equiv \binom{m+r}{r} \mod p, \tag{4.12}$$

$$\beta_{p^r+1} \equiv \binom{m+r+1}{r+1} \mod p. \tag{4.13}$$

To adapt these to the general n in $\xi_n^{(m)}$, r is replaced in (4.8)–(4.11) by an appropriate expression in the scale of p, when the new congruences become difference congruences of the same general type as those already solved. For example, to adapt (4.9) to β_n, where

$$n = r_j p^j + r_{j-1} p^{j-1} + \cdots + r_1 p + s, \qquad 0 < r_j < p,$$

$$0 \leqq r_{j-i} < p, \qquad i = 1, \cdots, j-1, \qquad j > 1,$$

we replace r by $r_j p^{j-1} + r_{j-1} p^{j-2} + \cdots + r_1$, and obtain

$$(\beta^{p^j} - \beta^{p^{j-1}})^{r_j} (\beta^{p^{j-1}} - \beta^{p^{j-2}})^{r_{j-1}} \cdots (\beta^p - \beta)^{r_1} (\beta)_s \equiv \alpha_n \mod p.$$

The highest (symbolic) power of β on the left is β^n, and hence

$$\beta_n - \alpha_n \equiv L(\beta) \mod p,$$

where $L(\beta)$ is a linear function of β's with suffixes $< n$ expressed in the scale of p. The last is a difference congruence for $\xi_n^{(m)}$, m being the variable, whose solution is

$$\xi_n^{(m)} \equiv \sum_{s=0}^{m} L(\xi^{(s)}) \mod p.$$

If the expressions as integral polynomials in m of the residues of the several terms of $\xi_i^{(s)}$ in L are known, the sum can be obtained by using Bernoulli polynomials, as before; if any of the residues of the $\xi_i^{(s)}$ are not yet known explicitly, the processes are repeated on them, and so on, using (4.8)–(4.13) and any results already found. The effect of the initial substitution for r is to precipitate all possible reductions of the crude degree by Fermat's theorem and the properties of binomial coefficients in one reduction. At any step there are usually alternative ways of proceeding, according to which one of (4.8)–(4.13) is used. The determination of the desired residue is effected by the solution of a chain of

difference congruences, the solution of a particular congruence in the chain being used in that of some subsequent congruence. We shall illustrate the process for $n = 3p^2$.

To obtain the residue of $\xi_{3p^2}^{(m)} \bmod p$, we take $r = 3p$ in (4.11), to produce the desired suffix $3p^2$. The left is then $(\beta^{p^2} - \beta^p)^3$, and hence

$$\beta_{3p^2} - \alpha_{3p^2} \equiv 3\beta_{2p^2+p} - 3\beta_{p^2+2p} + \beta_{3p} \bmod p.$$

The first and second terms on the right are not yet known; their suffixes direct us to take $r = 2p + 1$, $r = p + 2$ in (4.11). The third term is known from the third of (4.3). Proceeding thus we get the required chain:

$$(4.14)\quad \beta_{p^2+2} - \alpha_{p^2+2} \equiv \alpha_2\beta_{p^2} + \alpha_{p^2+1} + \beta_{p^2+1} \bmod p,$$

$$(4.15)\quad \beta_{p^2+p} - \alpha_{p^2+p} \equiv \beta_{p^2+1} + \beta_{2p} - \beta_{p+1} \bmod p,$$

$$(4.16)\quad \beta_{p^2+p+1} - \alpha_{p^2+p+1} \equiv \alpha_{p+1}\beta_{p^2} + \alpha_{p^2+1}\beta_p + \beta_{p^2+p} \bmod p,$$

$$(4.17)\quad \beta_{2p^2} - \alpha_{2p^2} \equiv 2\beta_{p^2+p} - \beta_{2p} \bmod p,$$

$$(4.18)\quad \beta_{2p^2+1} - \alpha_{2p^2+1} \equiv 2\alpha_{p^2+1}\beta_{p^2} + \beta_{2p^2} \bmod p,$$

$$(4.19)\quad \beta_{p^2+2p} - \alpha_{p^2+2p} \equiv 2\beta_{p^2+p+1} - \beta_{p^2+2} + \beta_{3p} - 2\beta_{2p+1} + \beta_{p+2} \bmod p,$$

$$(4.20)\quad \beta_{2p^2+p} - \alpha_{2p^2+p} \equiv 2\beta_{p^2+2p} + \beta_{2p^2+1} - 2\beta_{p^2+p+1} - \beta_{3p} + \beta_{2p+1} \bmod p,$$

$$(4.21)\quad \beta_{3p^2} - \alpha_{3p^2} \equiv 3\beta_{2p^2+p} - 3\beta_{p^2+2p} + \beta_{3p}.$$

By using (2.10), (4.1)–(4.6) when necessary, these eight congruences can be solved successively. They are obtained from (4.10), (4.11) by the following substitutions: (4.14), (4.16), (4.18) come from (4.10) by $(r, s) = (p, 1), (p, p), (2p, 0)$; (4.15), (4.17), (4.19), (4.20), (4.21) come from (4.11) by $r = p + 1, 2p, p + 2, 2p + 1, 3p$ respectively. It will be sufficient to indicate the solution of (4.14), and merely give the results for the others. From (2.10), (4.5), (4.6), (4.14) we get

$$\xi_{p^2+2}^{(m)} - \xi_{p^2+2}^{(m-1)} \equiv m\binom{m+2}{2} + \binom{m+2}{3} + \binom{m+3}{3} \bmod p;$$

$$\xi_{p^2+2}^{(m)} \equiv \tfrac{1}{6}\sum_{s=0}^{m}(s+1)(s+2)(2s+3) \bmod p;$$

$$(4.22)\qquad \xi_{p^2+2}^{(m)} \equiv \tfrac{1}{24}(m+1)(m+2)(m+3)(5m+4) \bmod p.$$

Proceeding from this by mathematical induction we find

(4.23) $\xi_{p^2+n}^{(m)} \equiv A_n(m) \bmod p$, *where* $A_n(m)$ *is an integral polynomial in* m *of degree* $n + 2$, *algebraically divisible by* $m + 1$.

The last is generalized at once by mathematical induction in an obvious way:

(4.24) $\xi_{p^r+s}^{(m)} \equiv C_{r,s}(m) \bmod p$, *where* $C_{r,s}(m)$ *is an integral polynomial in* m *of degree* $r + s$.

The solutions of (4.15)–(4.21) are obtained as outlined, and are as follows, with the additional information in (4.26), (4.27).

$$(4.25)\qquad \xi^{(m)}_{p^2+p} \equiv \tfrac{1}{24}(m+1)(m+2)(m+3)(3m+4) \mod p.$$

(4.26) $\xi^{(m)}_{p^r+p} \equiv C_r(m) \mod p$, *where* $C_r(m)$ *is an integral polynomial in* m *of degree* $r+2$;

$$(4.27)\qquad C_3(m) \equiv \tfrac{1}{120}(m+1)(m+2)(m+3)(m+4)(4m+5).$$

$$(4.28)\qquad \xi^{(m)}_{p^2+p+1} \equiv \tfrac{1}{120}(m+1)(m+2)(m+3)(13m^2+32m+20) \mod p.$$

$$(4.29)\qquad \xi^{(m)}_{2p^2} \equiv \tfrac{1}{60}(m+1)(m+2)(m+3)(3m^2+12m+10) \mod p.$$

Incidentally in the proof of the last the following curious relations appear,

$$\xi^{(m)}_{p^2} \equiv \xi^{(m)}_{p+1}, \qquad \xi^{(m)}_{2p^2} - \xi^{(m-1)}_{2p^2} \equiv [\xi^{(m)}_{p^2}]^2 \mod p.$$

$$(4.30)\qquad \xi^{(m)}_{2p^2+1} \equiv \tfrac{1}{360}(m+1)(m+2)(m+3)(13m^3+69m^2+113m+60) \mod p.$$

$$(4.31)\qquad \xi^{(m)}_{p^2+2p} \equiv \tfrac{1}{360}(m+1)(m+2)(m+3)(25m^3+93m^2+122m+60) \mod p.$$

$$(4.32)\qquad \xi^{(m)}_{2p^2+p} \equiv \tfrac{1}{360}(m+1)(m+2)(9m^5+85m^4+309m^3 + 548m^2+489m+180) \mod p.$$

$$(4.33)\qquad \xi^{(m)}_{3p^2} \equiv \tfrac{1}{6720}(m+1)(m+2)(m+3)F(m) \mod p,$$

$$F(m) \equiv 63m^5+570m^4+2005m^3+3498m^2+3104m+1120.$$

At any link of such a chain a cumulative check on all preceding work is obtained by taking $m = 1$ in the last link, and calculating $\xi^{(1)}_m$ directly, as can easily be done by repeated use of

$$(4.34)\qquad \xi^{(1)}_{rp+n} \equiv \sum_{s=0}^{r} \binom{r}{s} \xi^{(1)}_{n+s} \mod p,$$

which follows by successive substitutions of $p+n$ for n in the initial congruence

$$\xi^{(1)}_{p+n} \equiv \xi^{(1)}_n + \xi^{(1)}_{n+1} \mod p.$$

All of the preceding work was checked in this way. The $\xi^{(m)}_n$ increase so rapidly with m, n that extensive numerical checks are out of the question; however, where feasible, such checks were also used. The general conclusion, established by mathematical induction, is

(4.35) *If* $n = ap^\alpha + bp^\beta + \cdots + cp^\gamma + d$,

$$a, \alpha, b, \beta, \cdots, c, \gamma > 0, \qquad d \geqq 0, \qquad \textit{then } \xi^{(m)}_n \equiv R(m), \mod p,$$

where $R(m)$ *is an integral polynomial in* m *whose degree depends only on* a, α, $b, \beta, \cdots, c, \gamma, d$.

5. The residue of $\xi_n^{(m)}$ mod p is readily determined for a fixed n by preceding theorems when the residue of m mod p is known. From (1.15), (1.16), (4.35) we have

(5.1) *If* $m \equiv s \bmod p$, *and if* p *is not an exceptional prime for* $R(m)$ *in* (4.35), *then* $\xi_n^{(m)} \equiv R(s) \bmod p$.

The exceptional primes in any case require separate consideration, as in the following instance, which is typical of all. From the second of (4.3) we have

$$\xi_{2p}^{(m)} \equiv \tfrac{1}{6}(m+1)(m+2)(2m+3) \mod p,$$

in which the exceptional primes are 2, 3 (the prime divisors of 6). The corresponding values of m are $2r$, $2r+1$ for $p=2$, and $3r$, $3r+1$, $3r+2$ for $p=3$, and the residues of $\xi_{2p}^{(m)}$ for these values are required. In the calculations, all polynomials with rational coefficients take integral values $\geqq 0$ for integral values $\geqq 0$ of the variables, by the manner in which they are obtained from the given integral polynomial.

Reducing the coefficients mod 6 we have

$$\tfrac{1}{6}(m+1)(m+2)(2m+3) \equiv (m+1)^2 + \tfrac{1}{6}m(m+1)(2m+1);$$

and hence, for $p=2$,

$$\xi_4^{(m)} \equiv (m+1)^2 + \tfrac{1}{6}m(m+1)(2m+1) \mod 2.$$

Taking $m=2r$, $m=2r+1$, and reducing, we find

$$\xi_4^{(2r)} \equiv \xi_4^{(2r+1)} \equiv 1 + \tfrac{1}{3}r(2r^2+1) \mod 2.$$

In these we take $r=3t$, $3t+1$, $3t+2$, and reduce mod 2. Finally, then, we have the results stated for $\xi_4^{(m)}$ in (5.2).

Similarly we obtain

$$\xi_6^{(3r)} \equiv r^2 + 1 + \tfrac{1}{2}r(r+1) \mod 3,$$

$$\xi_6^{(3r+1)} \equiv \xi_6^{(3r+2)} \equiv r^2 + 2 + \tfrac{1}{2}r(r+1) \mod 3.$$

Taking $r=2t$, $2t+1$, and $t=3u$, $3u+1$, $3u+2$ in these, we find the results stated for $\xi_6^{(m)}$ in (5.2).

(5.2) $m \equiv r \bmod p$, $\quad p>3$, $\quad \xi_{2p}^{(m)} \equiv \tfrac{1}{6}(r+1)(r+2)(2r+3) \mod p$;

$m \equiv 2, 3, 6, 7, 10, 11 \bmod 12$, $\xi_4^{(m)} \equiv 0 \bmod 2$,

$m \equiv 0, 1, 4, 5, 8, 9 \bmod 12$, $\xi_4^{(m)} \equiv 1 \bmod 2$;

$m \equiv 3, 7, 8, 12, 16, 17 \bmod 18$, $\xi_6^{(m)} \equiv 0 \bmod 3$,

$m \equiv 0, 4, 5, 9, 13, 14 \bmod 18$, $\xi_6^{(m)} \equiv 1 \bmod 3$,

$m \equiv 1, 2, 6, 10, 11, 15 \bmod 18$, $\xi_6^{(m)} \equiv 2 \bmod 3$.

The information in (5.2) refers to $\xi_n^{(m)}$ mod p, where $n=2p$, that is, to the residues (or period) mod p of the $2p^{\text{th}}$ columns in the table described in §1.

We now seek all such information implied in results of the kind in §4. Since the first of (5.2) is complete for $p > 3$, we need attend only to $n = 4, 6$, and find the primes p for which 4, 6 are expressible as polynomials in p with non-negative integer coefficients. The only new possibilities are $p = 3$, $p + 1 = 4$; $p = 5$, $p + 1 = 6$ for which the first of (4.4) is applicable, and $p = 2$, $p^2 + p = 6$, an instance of (4.25). Proceeding as above, we find

$$(5.3)\quad m \equiv 1, 2, 4, 5 \mod 6,\ \xi_4^{(m)} \equiv 0 \mod 3,$$

$$m \equiv 0, 3 \mod 6,\ \xi_4^{(m)} \equiv 1 \mod 3.$$

$$(5.4)\quad m \equiv 3, 4, 8, 9 \mod 10,\ \xi_6^{(m)} \equiv 0 \mod 5,$$

$$m \equiv 0, 2, 5, 7 \mod 10,\ \xi_6^{(m)} \equiv 1 \mod 5,$$

$$m \equiv 1, 6 \mod 10,\ \xi_6^{(m)} \equiv 3 \mod 5.$$

$$(5.5)\quad m \equiv 4, 5, 6, 7, 12, 13, 14, 15, 20, 21, 22, 23 \mod 24,\ \xi_6^{(m)} \equiv 0 \mod 2,$$

$$m \equiv 0, 1, 2, 3, 8, 9, 10, 11, 16, 17, 18, 19 \mod 24,\ \xi_6^{(m)} \equiv 1 \mod 2.$$

Thus far only prime moduli p have been considered. Results such as the foregoing also give information concerning the residues for certain composite moduli. For if $p_i (i = 1, \cdots, r + 1)$ are distinct primes such that

$$\xi_n^{(m)} \equiv 0 \mod p_i \qquad (i = 1, \cdots, r + 1),$$

then

$$\xi_n^{(m)} \equiv 0 \mod p_1 \cdots p_{r+1}.$$

Thus, from the fourth of (5.2), the first of (5.4) and (5.5) we obtain the following, on noting that the L. C. M. of 18, 10 is 90, etc.:

$$(5.6)\quad m \equiv 7, 12, 21, 30, 39, 44, 52, 53, 61, 62, 70, 71 \mod 72,\ \xi_6^{(m)} \equiv 0 \mod 6;$$

$$m \equiv 4, 13, 14, 23, 28, 29, 38, 39, 44, 53, 54, 63, 68, 69, 78, 79, 84, 93, 94, 103, 108, 109, 118, 119 \mod 120,\ \xi_6^{(m)} \equiv 0 \mod 10;$$

$$m \equiv 3, 8, 34, 39, 43, 44, 48, 53, 79, 84, 88, 89 \mod 90,\ \xi_6^{(m)} \equiv 0 \mod 15;$$

$$m \equiv 39, 44, 53, 79, 84, 93, 124, 133, 134, 143, 174, 183, 188, 214, 223, 228, 268, 269, 278, 309, 318, 349, 358, 359 \mod 360,\ \xi_6^{(m)} \equiv 0 \mod 30.$$

A similar analysis may be made, in specific instances, for theorems of the kind (5.1). The implied restriction is necessary, as will appear in a moment,

because there does not seem to be much known concerning the forms of the prime divisors of polynomials in m of degree > 2 with integer coefficients, as m runs through all positive integers. Let n in (5.1) be expressed in the scale of the prime p_α, and choose $m_\alpha \equiv s_\alpha \bmod p_\alpha$, $\alpha = 1, \cdots, a$, $a > 1$, so that

$$\xi_n^{(m_\alpha)} \equiv R_\alpha(s_\alpha) \bmod p_\alpha, \qquad \alpha = 1, \cdots, a,$$

where p_α is not an exceptional prime for $R_\alpha(m_\alpha)$. If now $p_1, \cdots, p_a$ are all distinct, and $s_1, \cdots, s_a$ are such that

$$R_\alpha(s_\alpha) \equiv 0 \bmod p_\alpha, \qquad \alpha = 1, \cdots, a,$$

we can find m such that

$$\xi_n^{(m)} \equiv 0 \bmod p_1 \cdots p_a.$$

6. In what was called algebra C of one-rowed matrices or sequences,[4] reciprocals were defined as follows. Let θ, ϕ be the umbrae of $(\theta_0, \theta_1, \cdots)$, $(\phi_0, \phi_1, \cdots)$. Then θ, ϕ are called reciprocals of one another if

$$(\theta + \phi)^0 = 1, \qquad (\theta + \phi)^{n+1} = 0,$$

from which it follows that

$$\exp \theta x. \quad \exp \phi x = \exp\,(\theta + \phi)x = 1.$$

Also, if one of the sequences $\theta_0, \theta_1, \cdots, \phi_0, \phi_1, \cdots$ consists of integers, so does the other. If one of the sequences consists of positive integers, the other will contain zero or negative integers, or both.

The reciprocal $\eta^{(m)}$ of $\xi^{(m)}$ has arithmetical properties similar to those of $\xi^{(m)}$. For, from

$$\exp \xi^{(m)} x. \quad \exp \eta^{(m)} x = 1$$

and previous results we obtain

$$\eta_0^{(m)} = 1, \qquad \eta_n^{(0)} = (-1)^n; \tag{6.1}$$

$$\eta_{n+1}^{(m+1)} = \eta^{(m)}(\eta^{(m+1)} - \eta^{(m)})^n; \tag{6.2}$$

$$(\eta^{(m)} + \xi^{(m)})^{n+1} = 0; \tag{6.3}$$

$$\eta_n^{(m+1)} = \sum_{s=1}^{n} \eta_s^{(m)} \frac{\Delta^s 0^n}{s!}, \qquad n > 0; \tag{6.4}$$

$$(\eta^{(m+1)})_n = \eta_n^{(m)}; \tag{6.5}$$

$$f(\eta^{(m+1)}) = f(0) + \sum_{s=1} \frac{\Delta^s f(0)}{s!} \eta_s^{(m)}. \tag{6.6}$$

Comparing (6.2), (6.4), (6.5), (6.6) with (2.1), (1.10), (2.3), (2.7), we see by inspection what theorems in §§2–5 remain true when ξ is replaced by η, and what slight changes must be made in the statements of the others due to the

difference in signs in (2.1), (6.2). From the congruences for $\xi_n^{(m)}$, duals for $\eta_n^{(m)}$ can be written down from (6.3) on noting, for example, that $\eta_p^{(m)} \equiv -\xi_p^{(m)} \bmod p$ from (6.3) for $n = p - 1$. The reciprocity between ξ, η is much closer than that between any of the classical sequences, say those of the Bernoulli and Euler numbers, and their corresponding reciprocals. This is traced to the part played by e^x in the definitions, and the fact that e^x is its own derivative. The structure of the rectangular arrays for the $\xi_n^{(m)}$, $\eta_n^{(m)}$ are the same by (1.10), (6.4).

CALIFORNIA INSTITUTE OF TECHNOLOGY.

REFERENCES

1. G. Dobiński, Grunert's Archiv, vol. 61 (1877), pp. 333–6. E. Cesaro, Nouvelles Annales de Math., vol. 4 (1885), p. 39. M. D'Ocagne; Am. Jr. Math., vol. 9 (1887), pp. 370–2. F. Anderegg, Am. Math. Monthly, vol. 8 (1901), p. 54, vol. 9, pp. 11–13. P. Epstein, Archiv. der Math. u. Physik, vol. 8 (1904–5), pp. 329–30; A. Krug, ibid., vol. 9 (1905), pp. 189–91. [The generating function on p. 30 should be exp $(e^x - 1)$, not exp $(e^x + x - 1)$.] A. C. Aitken, Math. Notes (Edinburgh), Nr. 28, 1933, pp. 18–23. J. Touchard, Ann. Soc. Sci. Bruxelles, A 53, (1933), pp. 21–31. E. T. Bell, Am. Math. Monthly, vol. 41 (1934), pp. 411–19; Annals of Math., vol. 35 (1934), pp. 264–5, 267. For the Euler-Segner sequence, early references in E. Czuber, *Ber. üb. Wahrscheinlichkeitstheorie*, D. M. Ver. vol. 7 (1899), also E. Netto, *Kombinatorik*, p. 170. See also papers by Liouville, Lamé, Rodrigues, Catalan, Terquem in Liouville's Journal (1st series) vols. 3–6. Recent on same: A. Whitworth, *Choice and Chance*; E. Lucas, *Th. des Nombres*, p. 68; J. H. M. Wedderburn, Annals of Math., vol. 24 (1922–3), pp. 121–140 (the numbers occur incidentally); I. H. M. Etherington, Math. Gazette, vol. 21 (1937), pp. 36–9.

2. C. Jordan, Tôhoku Math. Jr., 37 (1933), pp. 255–283. (For the Stirling numbers).

3. E. T. Bell, *Algebraic Arithmetic*, 1927.

4. E. T. Bell, Trans. Am. Math. Soc., 25 (1923), pp. 255–83.

ANNALS OF MATHEMATICS
Vol. 39, No. 3, July, 1938

STRUCTURE RESIDUATION

BY MORGAN WARD

(Received Dec. 27, 1937)

I. INTRODUCTION

1. Given any two elements A and B of a structure[1] Σ, we define the residual of B with respect to A relative to cross-cut as an element $R = A : B$ of Σ with the properties

(1.1) $A \supset [A : B, B]$,

(1.2) $A \supset [X, B]$ implies that $A : B \supset X$ for any element X of Σ.

The residual $S = A - B$ of B with respect to A relative to union is defined dualistically by

$$(A - B, B) \supset A,$$

$(X, B) \supset A$ implies that $X \supset A - B$ for any element X of Σ.

The properties of these two operations may be summarized as follows: We define a structure as *residually closed* (relative to cross-cut) if any two elements in it have a residual with the properties (1.1), (1.2). *Every residually closed structure is distributive.*[1] A structure will be said to be *distributively closed*[2] (relative to cross-cut) if given any two sets Θ and Φ of elements of Σ, and the set Γ of all cross-cuts of elements of Θ with elements of Φ, the union of Γ is the cross-cut of the unions of Θ and Φ.

Every distributively closed structure is residually closed.[3] Sufficient conditions that a structure Σ be distributively closed are (i) Σ contains an all element O_0; (ii) Σ is distributive; (iii) the ascending chain condition holds in Σ. Thus every finite distributive structure is residually closed. Similar results hold by duality for residuation with respect to union.

2. The results of a recent paper (Ward [1]) in which I have considered residuation in a structure over which a multiplication is defined which is distributive with respect to union apply to the present case on identifying multiplication with cross-cut. The assumption $(A, B) = A : (A : B)$ analyzed there is a

---

[1] We use the terminology of Ore's recent memoir in these Annals (Ore, [1]), save for the substitution of the term "distributive structure" for Ore's "arithmetic structure," and the interchange of the symbols $(\cdots)$ and $[\cdots]$.

[2] MacNeille [1] uses the term "lattice with completely distributive products."

[3] Compare MacNeille [1] lemma 7.6. MacNeille's "product complement" A' is our "residual of A with respect to E_0."

necessary and sufficient condition that a structure residually closed with respect to cross-cut and containing a unit element E_0 may be a Boolean algebra.

If Σ is a distributive structure containing an all element in which the ascending chain condition holds, then the general decomposition theorems for Dedekind structures become almost trivial (Ore [1] page 415); every element of Σ save O_0 has a unique representation as the cross-cut of a finite number of prime elements none of which divides any other. A very interesting situation arises however if the accompanying structure residuation has any one of the three equivalent properties[4]

$$A : [B, C] = (A : B, A : C),\ (B, C) : A = (B : A, C : A),\ (A : B, B : A) = O_0 .$$

In this case, the primes in the canonical representation of any element as a cross-cut may be chosen relatively prime in pairs. The primes themselves are not necessarily powers of irreducible elements as in common arithmetic, but they may be associated with such powers in a relatively simple manner.

I propose here the name "semi-arithmetical structure (or lattice)" for these systems. A simple example is the set of numbers 1, 2, 3, 6 and 12 with $[\cdots]$ and $(\cdots)$ the L.C.M. and G.C.D. operations. We may remark that *every finite distributive structure may be represented as a finite set of positive integers closed under* L.C.M. *and* G.C.D. I have constructed a proof of this fact by induction; but it is also an immediate consequence of the recent result of Mac Neille's [1] allowing us to imbed any distributive structure in a Boolean algebra. Conversely, it implies Mac Neille's theorem for the case of a finite structure.

II. Properties of Residually Closed Structures

3. We define a structure following Ore as a system Σ of elements $A, B, C, \cdots$ over which there is a well defined division relation $X \supset Y$ such that

Postulate I. *$A \supset A$; $A \supset B$ and $B \supset C$ imply $A \supset C$.*

Postulate II. *For each pair of elements A, B of Σ there exist elements D and M such that: $D \supset A$, $D \supset B$; $S \supset A$, $S \supset B$ implies $S \supset D$. $A \supset M$, $B \supset M$; $A \supset T$, $B \supset T$ implies $M \supset T$.*

If $A \supset B$, $B \supset A$ we write $A = B$. We write $M = [A, B]$ $D = (A, B)$ reversing Ore's usage.

We shall now examine the consequences of assuming

Postulate III. (a). *Σ is residually closed relative to cross-cut; or* (b) *Σ is residually closed relative to union.*

Assume III a. Then taking $B = A$ in (1.1), we see that Σ contains an all element $O_0 = A : A$ dividing every other element. Similarly on assuming II b, we see that Σ contains a unit element $E_0 = A - A$ divisible by every other element. Since all properties of residuation relative to union may be obtained

[4] The first two equalities are the third and fourth "distributive laws for residuation" analyzed at length in Ward [1].

by a similar dualizing, we shall confine ourselves to stating results only for residuation relative to cross-cut.

We list here for reference some elementary properties of residuation. We use $\rightarrow$ for "implies that", $\cdots + \cdots$ for "both $\cdots$ and $\cdots$" and $\backsim$ for "implies and is implied by."

(i) $A : B \supset A.$

(ii) $A \supset B \rightarrow A : C \supset B : C + C : B \supset C : A.$

(iii) $A = B \rightarrow A : C = B : C + C : A = C : B.$

(iv) $A \supset B \backsim A : B = O_0.$

(v) $A : A = O_0 + A : O_0 = A.$

(vi) $A : B \supset C \backsim A \supset [B, C].$

(vii) $A : (A : B) \supset (A, B).$

(viii) $A : [B, C] = (A : B) : C = (A : C) : B.$

(ix) $A : B = (A : B) : B.$

(x) $[A, B] : B = A : B.$

(xi) $A : B = R \rightarrow R : B = R.$

(xii) If $A : B = A$, then $A \supset [X, B] \backsim A \supset X.$

(xiii) $R = A : B + S = A : R \rightarrow R = A : S.$

(xiv) If $B^{(1)} = B$, $B^{(n+1)} = A : B^{(n)}$, then

$B^{(2n)} = B^{(2)}$, $B^{(2n+1)} = B^{(3)}$, $(n = 1, 2, 3, \cdots).$

(xv) $A : B = A : (A, B).$

(xvi) $(A, B) = O_0 \rightarrow A : B = A.$

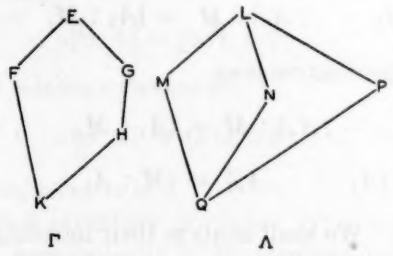
These rules all readily follow from the definition (1.1), (1.2) and postulates I, II, III a. Consider for example (xv). Since $(A, B) \supset B$, $A : B \supset A : (A,B)$ by (ii). By (vii), $A : (A : B) \supset (A, B)$. Therefore by (ii) again, $A : (A, B) \supset A : \{A : (A : B)\}$. But by (xiii), $A : \{A : (A : B)\} = A : B$. Hence $A : (A, B) \supset A : B$ by (iii), giving (xv).

THEOREM 3.1. *If Σ is residually closed relative to cross-cut, then Σ is a distributive structure.*

PROOF. All structures of order less than five are easily seen to be distributive. If Σ is not a Dedekind structure, it is known (Dedekind [1] p. 255) that Σ contains a sub-structure Γ of order five which is not a Dedekind structure. Similarly, if Σ is not distributive but is a Dedekind structure, it contains a substructure Λ of order five which is not distributive. (Birkhoff [1] p. 617.) The types of these sub-structures are well known; their lattice diagrams are given in the illustration (see next page). (Klein [1] pp. 222–223).

Suppose that Σ is not a Dedekind structure. Consider the residual $H : G$

of the elements G and H of the sub-structure Γ. Since $H \supset [F, G]$, $H : G \supset F$ by (1.2). Also $H : G \supset H$ by rule (i). Therefore $H : G \supset (H, F)$ or $H : G \supset E$. Then by rule (vi), $H \supset [G, E]$ or $H \supset G$ which is false.



Γ Λ

The structure Σ must therefore be Dedekindian. If it is non-distributive, consider the residual $M : N$ of the elements M and N in the sub-structure Λ. Since $M \supset [N, P]$, $M : N \supset P$ by (1.2). Also $M : N \supset M$ by rule (i). Therefore $M : N \supset (M, P)$ or $M : N \supset L$. Then by rule (vi), $M \supset [N, L]$ or $M \supset N$ which is false.

Hence every structure of order greater than four which is residually closed must be distributive, and the proof is complete.

4. Theorem 3.1 allows us to apply all the results obtained in Ward [1] for residuation in a structure closed with respect to multiplication. For on identifying cross-cut with the multiplication defined there, the three conditions given for a multiplication $X \cdot Y$ are satisfied (Ward [1] section 3); namely

$A \cdot B$ is in Σ if A, B are in Σ; $A \cdot (B \cdot C) = (A \cdot B) \cdot C$; $A \cdot B = B \cdot A$.

$O_0 \cdot A = A$ for every A in Σ.

$A \cdot (B, C) = (A \cdot B, A \cdot C)$.

In particular, if when $B \supset A$ we define the quotient of A by B as an element A/B of Σ such that

$$(4.1) \qquad A = \left[B, \frac{A}{B}\right], \qquad A = [B, X] \text{ implies that } \frac{A}{B} \supset X,$$

it follows that if the quotient A/B exists it equals the residual $A : B$. But more is true; namely

THEOREM 4.1. *If $B \supset A$, the quotient A/B always exists and equals the residual $A : B$.*

PROOF. By (1.1), $A \supset [A : B, B]$. But $B \supset A$ and $A : B \supset A$ by rule (i). Hence $[A : B, B] \supset A$, $A = [A : B, B]$. Also if $A = [B, X]$, $A \supset [B, X]$ so that $A : B \supset X$ by (1.2). Therefore $A : B = A/B$ by the definition (5.1).

We obtain as a corollary the following important rule of manipulation[5]:

(xvii) $$B \supset A \backsim A = [A : B, B].$$

[5] We may observe that (xvii) shows that postulate D of Ward [1] is always satisfied in a residually closed structure.

We also have the two "distributive laws" (Ward [1], section III)

LI $$M : (A_1, A_2, \cdots, A_n) = [M : A_1, \cdots, M : A_n]$$

LII $$[A_1, A_2, \cdots, A_n] : M = [A_1 : M, \cdots, A_n : M].$$

The remaining two distributive laws

LIII $$(A_1, \cdots, A_n) : M = (A_1 : M, \cdots, A_n : M)$$

LIV $$M : [A_1, \cdots, A_n] = (M : A_1, \cdots, M : A_n)$$

are not generally valid. We shall analyze their meaning and validity when we come to discuss the arithmetical properties of residually closed structures.

A large portion of Ward [1] was devoted to analyzing the consequences of assuming*

POSTULATE C. *If $A \supset B$, there exists an element F such that $A = B : P$.*

This assumption was shown to be equivalent to assuming the rule

(4.2) $$(A, B) = A : (A : B).$$

In case Σ has a unit element E_0, this last rule implies that Σ is a Boolean algebra. For consider $A' = E_0 : A$ for any element A of Σ. We have $(A, A') = A' : (A' : A) = E_0 : A : \{(E_0 : A) : A\} = O_0$ by rules (iii), (iv) and (v). Also since $A \supset E_0$, $[A, A'] = E_0$ by rule (xvii). Finally $(A')' = E_0 : (E_0 : A) = (E_0, A) = A$. Thus A' is the negative of A. Conversely in a Boolean algebra, $A : B = (A, B')$ (Dilworth [1]), and (4.2) is easily seen to be satisfied.

III. DISTRIBUTIVELY CLOSED STRUCTURES

5. Let Σ^* denote the class of all sub-sets of elements of Σ. If Θ and Φ are any two such sub-sets, we define $[\Theta, \Phi]$ and (Θ, Φ) to be the least sub-sets containing respectively all elements $[T, F]$ and (T, F), T in Θ, F in Φ. The operations $[\cdots]$ and $(\cdots)$ thus defined over Σ^* are commutative and associative and if Σ contains an all element or a unit element so does Σ^*, but the resulting algebra is not a structure since the operations are obviously not indempotent.

If every element of Θ divides every element of Φ, we write $\Theta \supset \Phi$. If $\Phi \supset \Phi$, $[\Theta, \Phi] = \Phi$, $(\Theta, \Phi) = \Phi$ but not conversely. If Θ consists of a single element T, we write when convenient $\Phi \supset T$, $T \supset \Phi$, $[T, \Phi]$ or (T, Φ) for $\Phi \supset \Theta, \Theta \supset \Phi$, $[\Theta, \Phi]$ or (Θ, Φ).

The assumption that Σ is a closed structure (Ore [1] page 409) may then be formulated by saying that we have two operators u and k defined on Σ^* to Σ such that

$$u\Theta = U;\ U \supset \Theta;\ X \supset \Theta \text{ implies } X \supset U.$$

$$k\Theta = K;\ \Theta \supset K;\ \Theta \supset Y \text{ implies } K \supset Y.$$

6. Let Σ be any closed structure. Then Σ will be said to be *distributively closed (relative to cross-cut)* if

$$u[\Theta, \Phi] = [u\Theta, u\Phi]$$

and *distributively closed relative to union if*

$$k(\Theta, \Phi) = (k\Theta, k\Phi)$$

Here Θ, Φ denote any two sub-sets of Σ.

A distributively closed structure of either type is obviously distributive ("arithmetic" is Ore's terminology) in the ordinary sense. A distributively closed structure relative to both union and cross-cut is a $\bar{C}$ lattice in Garrett Birkhoff's terminology (Birkhoff [2]); conversely, it is easily shown that any $\bar{C}$ lattice is a distributively closed structure relative to both union and cross-cut.[6]

The assumption that Σ is distributively closed relative to cross-cut is equivalent to postulate B of Ward [1] on identifying the multiplication $X \cdot Y$ with $[X, Y]$. We accordingly obtain by the argument given in Ward [1] section 4

THEOREM 6.1. *Every structure distributively closed relative to cross-cut is residually closed relative to cross-cut.*

THEOREM 6.2. *Let Σ be a distributive structure containing an all element O_0 in which the ascending chain condition holds. Then Σ is distributively closed relative to cross-cut.*

PROOF. The hypotheses of the theorem imply that Σ is closed relative to union in the ordinary sense. (Ore [1] §2.) Let Θ and Φ be any two sub-sets of Σ with elements $\cdots T_r \cdots ; \cdots F_s \cdots$ and let $\Gamma = [\Theta, \Phi]$. Then $T = u\Theta$, $F = u\Phi$ and $G = u\Gamma$ all exist. We are to prove that $G = [T, F]$. Now (Ore [1]§2) Θ contains k elments $T_1, \cdots, T_k$ such that $T = (T_1, \cdots, T_k)$. Similarly $F = (F_1, \cdots F_l)$ where the F_i are in Φ. Hence since Σ is distributive, $[T, F] = ([T_1, F_1], \cdots [T_k, F_l])$,

Also $G = ([T_{r_1}, F_{s_1}], \cdots, [T_{r_m}, F_{s_m}])$ for Γ is made up of all distinct elements $[T_r, F_s]$. Hence since $T \supset \Theta$, $F \supset \Phi$, $[T, F] \supset [T_{r_i}, F_{s_i}]$ or $[T, F] \supset G$. Also since $G \supset [T_r, F_s]$, $G \supset ([T_1, F_1], \cdots [T_k, F_l])$ or $G \supset [T, F]$. Thus $G = [T, F]$.

COROLLARY. *Every distributive structure of finite order is residually closed.*

IV. Arithmetical Properties of Residually Closed Structure

7. Let Σ be any structure. An element $Q \neq O_0$ of Σ is said to be: (i) *irreducible* if $X \supset Q$ implies $X = O_0$ or $X = Q$; (ii) *prime* if $Q \supset [X, Y]$ implies $Q \supset X$ or $Q \supset Y$; (iii) *indecomposable* if $Q = [X, Y]$ implies $Q = X$ or $Q = Y$; (iv) a *power* if it has precisely one irreducible divisor. Every irreducible is

[6] MacNeille [1] calls such a structure a "completely distributive lattice," and proves the independence of distributive closure with respect to union and distributive closure with respect to cross-cut.

indecomposable, and in a distributive structure every irreducible is a prime. (Köthe [1].)

For the remainder of the paper we assume postulate III a and a chain condition:

POSTULATE IV. *The ascending chain condition holds in* Σ

Then as is well known every element A of $\Sigma \neq O_0$ admits of at least one decomposition:

$$A = [Q_1', Q_2', \cdots, Q_l'] \tag{7.1}$$

into a cross-cut of a finite number of indecomposable elements Q' of Σ.

By theorem 3.1, Σ is distributive and hence Dedekindian. Now Σ contains O_0 and hence by the ascending chain condition at least one irreducible P. If P' is any power of the irreducible P, we define the "multiplicity" of P' as the length of any principal chain (Ore [1] page 411) $P > \cdots > P'$ joining P and P' increased by unity. Then every power of P has a unique finite multiplicity, and P itself is of multiplicity unity. It is convenient to consider the all element O_0 as the unique power of P of multiplicity zero. The following lemmas are obvious from this definition:

LEMMA 7.1. *If R and S are both powers of P and $R \supset S$, $R \neq S$ then the multiplicity of R is less than the multiplicity of S.*

LEMMA 7.2. *The powers of any irreducible form a dense sub-structure of* Σ.

Now let A be any element of $\Sigma \neq O_0$. Then by the chain condition, A has at most a finite number of irreducible divisors $P_1, P_2, \cdots P_k$ and for each irreducible P_i there is a power P_i' of the highest positive multiplicity which divides A. Then the element $[P_1', P_2', \cdots, P_k']$ is a divisor of A which we call a *kernel* of A.

THEOREM 7.1. *Every element of* Σ *has exactly one kernel.*

PROOF. Assign to O_0 the kernel O_0. If $A \neq O_0$, A has at least one kernel by the argument above. Let P' and P'' be powers of the irreducible divisor P of A of maximum multiplicity m dividing A. It suffices to show that $P' = P''$. Let $P^* = [P', P'']$. Then $P^* \supset A$ and P^* is a power of P by lemma 7.2. If $P^* \neq P'$ then since $P' \supset P^*$, the multiplicity of P^* is greater than m by lemma 8.1. Hence $P^* = P'$. Similarly, $P^* = P''$, $P' = P''$.

We denote the kernel of any element A of Σ by θA. If $A = \theta A$, A will be said to be *regular*. Thus all powers are regular. Operations on kernels are governed by the following simple rules:

(7.1) $A \supset B$ implies that $\theta A \supset \theta B$.

(7.2) $\theta A = O_0$ if and only if $A = O_0$.

(7.3) $\theta(A, B) = (\theta A, \theta B)$, $\theta[A, B] = [\theta A, \theta B]$.

(7.4) $\theta\theta A = \theta A$.

The converse of rule (7.1) is generally false. It is obvious from (7.3) that we have:

THEOREM 7.2. *The set of all regular elements of Σ forms a sub-structure of Σ.*

THEOREM 7.3. *The set of all elements of Σ with the same kernel forms a sub-structure of Σ.*

THEOREM 7.4. *If A is regular and $B \supset A$, then B is regular.*

PROOF. We observe that if P' is any power of P and M any element of Σ, then it follows from lemma 7.2 that $P'' = (M, P')$ is also a power of P. Assume that $B \supset A$, A regular. Then

$$B = (B, A) = (B, \theta A) = (B, [P'_1, \cdots, P'_k])$$
$$= [(B, P'_1), \cdots (B, P'_k)] = [P''_1, \cdots P''_k].$$

Hence by rule (7.3) $\theta B = \theta[P''_1, \cdots P''_k] = [\theta P''_1, \cdots \theta P''_k] = [P''_1, \cdots, P''_k] = B$.

Thus the sub-structure of all kernels of Σ is dense over Σ. (Ore [1], page 429).

8. So far no direct use has been made of postulate III a. It is obvious that the sub-structure of all regular elements and the sub-structure of all powers of any irreducible element are residually closed since the structures are dense over Σ and $A : B \supset A$ by rule (i) for residuation. Let A be any indecomposable element of Σ, B any other element of Σ. Then if $A \supset B$, $A : B = O_0$ by rule (iv) and conversely. If $A \not\supset B$, $A : B = A : D$ where $D = (A, B)$ by rule (xv). Then by rule (xvii), $A = [A : D, D]$ Hence $A : D = A$. We therefore have

THEOREM 8.1. *If A is indecomposable, then $A : X = A$ or O_0 for every element X of Σ.*

THEOREM 8.2. *An element of Σ is a prime if and only if it is indecomposable.*

PROOF. Assume that A is indecomposable, $A \supset [B, C]$, $A \not\supset B$. We are to show that $A \supset C$. Since $A \not\supset B$, $A : B = A$ by theorem 8.1. But $A \supset [B, C]$ implies $A : B \supset C$ by (1.2) Hence $A \supset C$. The converse is trivial.

The following two theorems are immediate corollaries.

THEOREM 8.3. *Let A and B be any two elements of Σ with decompositions as in* (7.1) $A = [Q'_1, \cdots Q'_l]$, $B = [Q''_1, \cdots, Q''_k]$ *into cross-cuts of indecomposable elements. Then a necessary and sufficient condition that $B \supset A$ is that every Q'' should divide at least one Q'.*

THEOREM 8.4 "DECOMPOSITION THEOREM". *Let A be any element of $\Sigma \neq O_0$. Then A admits of a decomposition*

$$A = [Q_1, Q_2, \cdots, Q_l]$$

into a cross-cut of a finite number of indecomposable elements Q such that no Q_i divides a Q_j unless $i = j$. This decomposition is unique save for the order in which the factors Q are written.

Theorem 8.1 was recently proved in a different manner by Birkhoff (Birkhoff [3], page 452).

The following theorem will be useful subsequently.

THEOREM 3.5. *If A and B are any two primes of Σ, then one and only one of the following conditions hold:*

(i) $(A, B) = A$, (ii) $(A, B) = B$, (iii) $(A, B) = (A : B, B : A)$.

V. Semi-Arithmetical Structures

9. We shall now develop the consequences of assuming

Postulate V. *If M and N are any two elements of Σ, then*

$$(M : N, N : M) = O_0 .$$

A system satisfying postulates I, II, III a, IV and V will be called a "semi-arithmetical" structure.

Theorem 9.1. *In a residually closed structure, either the third or the fourth distributive law for residuation implies postulate V.*

Proof. We may take the third and fourth laws in the abbreviated forms

$$\text{L III} \qquad (A, B) : M = (A : M, B : M),$$

$$\text{L IV} \qquad M : [A, B] = (M : A, M : B),$$

the general forms in section 4 following by an easy induction. Assume L III. Then by rules (v) and (xv), $O_0 = (M, N) : (M, N) = (M : (M, N), N : (M, N) = (M : N, N : M)$. Assume L IV. Then by rules (v) and (x),

$$O_0 = [M, N] : [M, N] = ([M, N] : M, [M, N] : N) = (N : M, M : N).$$

Theorem 9.2. *In a semi-arithmetical structure, the powers of any irreducible form an ordered structure.*

Proof. If P' and P'' are powers of P, $P' : P''$ and $P'' : P'$ are both powers of P. But $(P' : P'', P'' : P') = O_0$. Hence either $P' : P'' = O_0$ or $P'' : P' = O_0$, so that by rule (iv), either $P' \supset P''$ or $P'' \supset P'$.

Theorem 9.3. *In a semi-arithmetical structure, all primes with the same kernel form an ordered structure.*

Proof. Let A and B be primes, $\theta A = \theta B = Q \neq O_0$. Then $Q \supset A, Q \supset B$ so that $Q \supset (A, B)$. Hence $(A, B) \neq O_0$. But by theorem 8.5 and postulate V, either $(A, B) = A$ or $(A, B) = B$ or $(A, B) = O_0$. Hence either $A \supset B$ or $B \supset A$.

Theorem 9.4. *In a semi-arithmetical structure, if A and B are primes and $\theta A \neq \theta B$, $A \supset B$ if and only if $\theta A \supset \theta B$.*

Proof. Assume that $\theta A \neq \theta B$ and $\theta A \supset \theta B$. Then $\theta(A, B) = (\theta A, \theta B) = \theta A \neq O_0$. But $\theta(A, B) \supset (A, B)$. Hence $(A, B) \neq O_0$. Therefore as in theorem 9.3 either $A \supset B$ or $B \supset A$. If $B \supset A$, $\theta B \supset \theta A$, $\theta B = \theta A$ contrary to hypothesis. Hence $A \supset B$. The converse is trivial.

Theorem 9.5. *Let T be any regular element of Σ. Then the structure Θ of all primes with the kernel T is dense over Σ.*

Proof. By theorem 9.3, Θ is ordered. Suppose that $A \supset X \supset B$, A, B in Θ, X in Σ. Then $\theta X = T$ by rule 7.1. By the decomposition theorem, $X = [C_1, C_2, \cdots C_t]$, $t \geqq 1$, where each C_i is a prime and divides no other C_j, $i \neq j$. Since $A \supset X$, A divides some C_i. Assume that $A \supset C_1$. Then since $C_1 \supset X$, $\theta C_1 = T$. But since $C_i \supset X$, $\theta C_i \supset \theta X$ or $\theta C_i \supset \theta C_1$ by rule (7.1). Thus by

theorem 9.4 either $C_t \supset C_1$ or $\theta C_t = T$. In the latter case, either $C_t \supset C_1$ or $C_1 \supset C_t$ by theorem 9.3. Thus in either case, $t = 1$ and X is a prime.

It follows that any set of prime elements with the same kernel form a principal chain.

We say that two elements R and S of Σ are relatively prime if $(R, S) = O_0$. The following theorem is now obvious:

THEOREM 9.6. *Every element $A \neq O_0$ of a semi-arithmetical structure may be uniquely represented save for order as the cross-cut of a finite number of prime elements which are relatively prime in pairs.*

THEOREM 9.7. *In a semi-arithmetical structure the third distributive law for residuation always holds; that is for any three elements A, B of Σ*

L III $$(A, B) : C = (A : C, B : C)$$

PROOF. The law is obviously true if either A or B is O_0. We next show it is true if A and B are primes Q and R. For then either (i) $(Q, R) = O_0$, or (ii) $(Q, R) = Q$ or (iii) $(Q, R) = R$. (i). If $(Q, R) = O_0$, $(Q, R) : C = O_0$. But then $(Q : C, R : C) = O_0$. For either $Q : C = O_0$, or $R : C = O_0$, or $Q : C = Q$ and $R : C = R$. Hence $(Q, R) : C = (Q : C, R : C)$ in this case. (ii). If $(Q, R) = Q$, then $(Q, R) : C = Q : C$. Hence if $Q : C = O_0$, $(Q, R) : C = (Q : C, R : C)$. If $Q : C = Q$, then $R : C = R$. For if $R : C = O_0$, $R \supset C$ But since $(Q, R) = Q$, $Q \supset R$, $Q \supset C$ and $Q : C = O_0$. Thus if $Q : C = Q$, $(Q : C, R : C) = (Q, R) = Q = (Q, R) : C$. In (iii) the proof is similar.

If A and B are not primes, let their canonical decompositions be

$$A = [\cdots, Q, \cdots], B = [\cdots, R, \cdots].$$

Then by the second distributive law for residuation

$$\begin{aligned}(A : C, B : C) &= ([\cdots, Q, \cdots] : C, [\cdots, R, \cdots] : C) \\ &= ([\cdots, Q : C, \cdots], [\cdots, R : C, \cdots]) \\ &= [\cdots, (Q : C, R : C), \cdots] \\ &= [\cdots, (Q, R) : C, \cdots] \\ &= [\cdots, (Q, R), \cdots] : C \\ &= ([\cdots, Q, \cdots], [\cdots, R, \cdots]) : C \\ &= (A, B) : C.\end{aligned}$$

In like manner we may prove

THEOREM 9.8. *In a semi-arithmetical structure, the fourth distributive law for residuation always holds.*

On combining theorems 9.1, 9.7 and 9.8, 6.1 and 6.2 we obtain

COROLLARY. *Either the third distributive law for residuation, the fourth distributive law for residuation or postulate V is a necessary and sufficient condition*

that a distributive structure containing an all element in which the ascending chain condition holds may be a semi-arithmetical structure.

Let us call a structure "arithmetical" if the only prime elements in it are powers of irreducibles. (Arithmetical structures have been exhaustively investigated by F. Klein who calls them "Sternverbande" (Klein [2])). We see then that an arithmetical structure is simply a semi-arithmetical structure in which all elements are regular, or in which the only primes are powers of irreducible elements.

REFERENCES

G. Birkhoff:
1. Bulletin Am. Math. Soc. vol 40 (1934), pp. 613–619.
2. Proc. Cambr. Phil. Soc. vol. 29 (1933), pp. 441–464.
3. Duke Math. Journal, vol. 3, Sept. (1937), pp. 443–454.

R. Dedekind: 1. Collected Works (1931), vol. II, paper 30, pp. 236–271.

R. P. Dilworth: 1. Bulletin Am. Math. Soc., vol. 44 (1938), pp. 262–268.

F. Klein:
1. Deutsche Math., vol. 2 (1937), pp. 216–241.
2. Math. Annalen., vol. 106 (1932), pp. 114–134.

G. Köthe: Deutsche Math. Verein, '37, pp. 125–144.

H. M. MacNeille: 1. Trans. Am. Math. Soc., vol. 42, November 1937, pp. 416–460.

O. Ore: 1. These Annals (2), vol. 36 (1935), pp. 406–432.

M. Ward: 1. Duke Math. Journal, vol. 3, December (1937), 627–636.

California Institute of Technology.

ANNALS OF MATHEMATICS
Vol. 39, No. 3, July, 1938

ON ONE PARAMETER GROUPS OF LINEAR TRANSFORMATIONS[1]

BY NELSON DUNFORD

(Received March 1, 1938)

In connection with the theorem of M. H. Stone[2] on the representation of one parameter groups of unitary transformations on Hilbert space J. von Neumann[3] has shown (without using any representation theorem) that if a group $U_t(-\infty < t < \infty)$ of unitary operators on Hilbert space H is measurable in the sense that $(U_t x, y)$ is measurable for each x, y in H, then U_t is necessarily continuous in the strong topology of the ring of bounded operators on Hilbert space, i.e.,

$$\lim_{t \to s} U_t x = U_s x, \qquad \text{for } x \text{ in } H.$$

The following question then naturally presents itself. What continuity properties must a group $T_t(-\infty < t < \infty)$ of linear mappings of a Banach space X into itself necessarily possess if it is assumed that for each x in X and y in $\bar{X}$ (the space conjugate to X) the numerical function $yT_t x$ is measurable in the sense of Lebesgue on $(-\infty < t < \infty)$? One answer to this question is that if for each x in X the set of values $T_t x(-\infty < t < \infty)$ is a separable set in X and if the norms $\| T_t \|$ are bounded on some interval $-a \leqq t \leqq a$ then the function T_t on $(-\infty, \infty)$ to $\mathfrak{X}$, the ring of linear mappings of X into itself, is continuous in the strong topology of $\mathfrak{X}$, i.e.,

$$\lim_{t \to s} T_t x = T_s x, \qquad \text{for } x \text{ in } X.$$

If the set of values of the function T_t on $(-\infty, \infty)$ to $\mathfrak{X}$ form a separable set in the uniform topology of $\mathfrak{X}$(i.e., in the metric $(T, T') = \| T - T' \| = \sup_{\|x\| \leq 1} \| Tx - T'x \|$) then

$$\lim_{t \to s} \| T_t - T_s \| = 0.$$

It seems desirable to word the theorem in terms of semi-groups so that the transformations T_t need not have inverses. By a *one parameter semi-group of linear transformations on* X will be meant a set of linear transformations

[1] Presented to the American Mathematical Society Feb. 26, 1938.

[2] M. H. Stone, *Linear Transformations in Hilbert Space* III, Proc. Nat. Acad. 16 (1930), pp. 172–175. See also F. Riesz, *Über Sätze von Stone und Bochner*, Acta Szeged, 6 (1933), pp. 184–198.

[3] J. von Neumann, *Über einen Satz von M. H. Stone*, Annals of Math., 33 (1932), pp. 567–573.

$T_t (0 < t < \infty)$ such that $T_t T_h = T_{t+h}$ for $0 < t, h < \infty$. In this case we only derive continuity on the right but it is readily seen from the argument given that in the case of groups the continuity is two-sided. A one parameter semi-group T_t of linear transformations on a Banach space X will be said to be *weakly measurable* if for every x in X and y in $\bar{X}$ the function $yT_t x$ is measurable on $(0, \infty)$, and it will be said to be *bounded* in case $\| T_t \|$ is bounded on some interval $(0, a)$.

THEOREM: *Let T_t be a weakly measurable and bounded one parameter semi-group of linear transformations on the Banach space X. If for each x in X the set $T_t x (0 < t < \infty)$ of image points is separable*[4] *then*

$$\lim_{s \to t+} T_s x = T_t x, \qquad x \text{ in } X, 0 < t < \infty.$$

If the set $T_t (0 < t < \infty)$ of points in $\mathfrak{X}$ is separable in the uniform topology of $\mathfrak{X}$ then

$$\lim_{s \to t+} \| T_s - T_t \| = 0, \qquad 0 < t < \infty.$$

The proof will be based on three properties of the integral of an absolutely integrable and measurable vector valued function of a real variable. A function $f(t)$ on $a \leqq t \leqq b$ to a Banach space X is said to be measurable if it is the limit almost everywhere of a sequence of step functions. Bochner[5] has developed a theory of Lebesgue integration of measurable vector valued functions $f(t)$ whose norm $\| f(t) \|$ is integrable, and has shown that such functions are continuous in the mean, i.e.,

$$\lim_{h \to 0} \int_a^b \| f(t + h) - f(t) \| \, dt = 0, \tag{1}$$

and that

$$\frac{d}{dt} \int_a^t f(s) \, ds = f(t) \qquad \text{almost everywhere.} \tag{2}$$

The integral $\int_a^b f(s) ds$ has its value in X and if V is any continuous linear mapping of X into a Banach space Z then[6] $Vf(t)$ is absolutely integrable and measurable and

$$V \int_a^b f(s) \, ds = \int_a^b Vf(s) \, ds. \tag{3}$$

---

[4] It is of course sufficient to assume instead, that X is a separable space.

[5] S. Bochner, *Integration von Funktionen deren Werte die Elemente eines Vektorraumes sind*, Fund. Math., 20 (1933) pp. 262-276.

[6] N. Dunford, *Integration and linear operations*, Trans. Amer. Math. Soc., 40 (1936), pp. 474-494, in particular Theorem 2.3.

Proceeding to the proof of the theorem we first note that if for a particular x in X the function $T_t x$ is continuous on the right at t_0 then it is continuous on the right at every point $s > t_0$. For let $t > s > t_0$, and $t \to s$, then

$$T_t x = T_{s-t_0} T_{t-s+t_0} x \to T_{s-t_0} T_{t_0} x = T_s x.$$

The fact that $\| T_t \|$ is bounded on some interval $(0, a)$ combined with the group property shows at once that it is bounded on any finite interval (a, b). Now Pettis[7] has shown that if a function $f(t)$ having its values in a separable subset of a Banach space X has the property that $yf(t)$ is measurable for every y in $\bar{X}$ then $f(t)$ is measurable. Thus for each x in X the abstract function $T_t x$ is measurable on $0 < t < \infty$. Since $\| T_t x \|$ is bounded in any finite interval the function $T_t x$ is an absolutely integrable and measurable function and so by (1) we have

$$\lim_{h \to 0} \int_a^b \| T_{t+h} x - T_t x \| \, dt = 0.$$

This shows upon using (3) that

$$\lim_{h=0} T_h \int_a^b T_t x \, dt = \lim_{h=0} \int_a^b T_{t+h} x \, dt = \int_a^b T_t x \, dt.$$

Thus

$$\lim_{s \to t+} T_s u = T_t u, \quad 0 < t < \infty,$$

for every u in X of the form

$$u = \int_a^b T_t x \, dt,$$

where a, b are arbitrary positive numbers and x is an arbitrary point in X. Now let x be fixed and define

$$u_n(t) = n \int_t^{t+1/n} T_s x \, ds.$$

Then for $0 < h \leqq 1$

$$\| T_h u_n(t) - T_h T_t x \| = \left\| T_h \left(n \int_t^{t+1/n} T_s x \, ds - T_t x \right) \right\|$$

$$\leqq M \left\| n \int_t^{t+1/n} T_s x \, ds - T_t x \right\|,$$

---

[7] B. J. Pettis, *Integration in Vector spaces*, to appear in the Trans. Amer. Math. Soc.

where M is a bound for $\| T_h \|$ on $(0, 1)$. Thus by (2) we have for each t in a set $E_0 \subset (0, 1)$ of measure one

$$\lim_{n=\infty} T_h u_n(t) = T_h T_t x \qquad \text{uniformly in } h.$$

As a function of h, $T_h u_n(t)$ is continuous on the right at every positive value of h and since the convergence as $n \to \infty$ is uniform with respect to h in $(0, 1)$ we must have for t in E_0

$$\lim_{h \to 0} T_{t+h} x = \lim_{h \to 0} T_h T_t x = T_t x.$$

But the measure of E_0 being one, t is as small as we please so that $T_t x$ is continuous on the right at every positive value of t.

Now suppose that $T_t (0 < t < \infty)$ is a separable set in the ring $\mathfrak{X}$ with its uniform topology, i.e., for each t, T_t is approachable in norm by a suitable chosen subsequence of some sequence $\{T_{t_n}\}$. The ring $\mathfrak{X}$ may now be regarded as a Banach space and we will first show that T_t on $(0, \infty)$ to $\mathfrak{X}$ is measurable. Let $x_{m,n,p}$ be a point in X with

$$\| x_{m,n,p} \| = 1, \qquad \| T_{t_m} x_{m,n,p} - T_{t_n} x_{m,n,p} \| > \| T_{t_m} - T_{t_n} \| - 1/2^p.$$

For a given t fix m so that

$$\| T_t - T_{t_m} \| < 1/2^p.$$

Then

$$\| T_t - T_{t_n} \| < 3/2^p + \| T_t x_{m,n,p} - T_{t_n} x_{m,n,p} \|,$$

which shows that for each $n = 1, 2, \cdots$

$$\sup_{p,i,j} \| T_t x_{i,j,p} - T_{t_n} x_{i,j,p} \| = \| T_t - T_{t_n} \|,$$

and thus since for each i, j, p, n the function $\| T_t x_{i,j,p} - T_{t_n} x_{i,j,p} \|$ is measurable on $(0, \infty)$ so is $\| T_t - T_{t_n} \|$. Now for every t there is an n such that

$$\| T_t - T_{t_n} \| < \epsilon. \tag{4}$$

Let E_n be the set of t for which (4) holds. Then E_n is measurable and the sets

$$E_1' = E_1, \qquad E_n' = E_n - \sum_{j=1}^{n-1} E_j$$

are measurable, disjoint and

$$(0, \infty) = \sum_{n=1}^{\infty} E_n'.$$

If we define

$$T_t^\epsilon = T_{t_n} \quad \text{on } E_n',$$

then T_t^ϵ is a denumerably valued measurable function on $(0, \infty)$ to $\mathfrak{X}$ and

$$\| T_t^\epsilon - T_t \| < \epsilon, \qquad 0 < t < \infty.$$

The function T_t on $(0, \infty)$ to the Banach space $\mathfrak{X}$ can therefore be approached uniformly by denumerably valued measurable functions and is thus measurable. Since $\| T_t \|$ is bounded on any finite interval we have the abstract function T_t measurable and summable on any finite interval. Any V in $\mathfrak{X}$ may be regarded as a linear operation on $\mathfrak{X}$ to $\mathfrak{X}$ and so we have, using equations (1), (2), (3), just as before

$$\lim_{h=0} T_h U = U$$

for every U of the form

$$U = \int_a^b T_s\, ds.$$

Thus if

$$U_n(t) = n \int_t^{t+1/n} T_s\, ds,$$

$\| T_h U_n(t) - T_h T_t \| \leqq M \left\| n \int_t^{t+1/n} T_s\, ds - T_t \right\| \to 0$ almost everywhere, and the convergence is uniform for $0 < h \leqq 1$. Now if $s > t_0$ and $s \to t_0$ then

$$T_s U_n(t) = T_{t_0} T_{s-t_0} U_n(t) \to T_{t_0} U_n(t),$$

i.e., for each n and t, $T_s U_n(t)$ is continuous on the right at each point. Fix t in the set where $\int_0^t T_s\, ds$ is differentiable and fix n so large that

$$\| T_h U_n(t) - T_h T_t \| < \epsilon/3 \quad 0 < h \leqq 1,$$

$$\left\| n \int_t^{t+1/n} T_s\, ds - T_t \right\| < \epsilon/3.$$

Then for $h > 0$ sufficiently small

$$\| T_h U_n(t) - U_n(t) \| < \epsilon/3,$$

so that for sufficiently small $h > 0$

$$\| T_{h+t} - T_t \| < \epsilon,$$

which, combined with the group property, proves that T_t is continuous on the right at each point in $(0, \infty)$.

It might be mentioned that the method used in proving T_t measurable could be used in proving that $T_t x$ is measurable, the fact used in the first part of the proof.

It might also be pointed out that in the case of groups the second conclusion follows immediately from the first by applying a well known theorem of Banach[8] which states that a closed everywhere defined distributive mapping of one complete separable group on to another is necessarily continuous.

YALE UNIVERSITY, NEW HAVEN, CONN.

[8] S. Banach, *Über metrische Gruppen*, Studia Math. III (1931) pp. 101–113, in particular Theorem 10 p. 112.

ANNALS OF MATHEMATICS
Vol. 39, No. 3, July, 1938

LINEAR OPERATIONS WHICH DEPEND ANALYTICALLY ON A PARAMETER

BY A. E. TAYLOR*

(Received December 13, 1937)

1. Introduction

It has long been customary in the theory of linear transformations—integral equations, linear equations in an infinite number of unknowns, etc.—to make use of certain notions from the theory of functions of a complex variable. In particular, the well known Neumann expansion of the resolvent is a formal power series whose coefficients, instead of being numbers, are iterates of a certain linear transformation.[1] These resemblances to the classical Cauchy-Weierstrass theory can be described accurately in terms of a theory of analytic functions whose values lie in a complex Banach space. The possibility of such a theory, derived in the usual manner from Cauchy's theorem and integral formula, was recognized by Wiener.[2] Still further generalizations, in which the independent variable may also range over a complex Banach space, have been considered by a number of people.[3] It therefore seems desirable to outline a general abstract theory of linear operations depending analytically on a parameter, which we shall here suppose to be a complex variable.

DEFINITION 1.1 *A function $f(\lambda)$ defined in an open set Δ of the complex plane, with values in a complex Banach space E, is called analytic in Δ if the limit*

$$\lim_{\lambda \to \lambda_0} \frac{f(\lambda) - f(\lambda_0)}{\lambda - \lambda_0}$$

exists at each point λ_0 of Δ. $f(\lambda)$ is said to be analytic at a point if it is analytic in some neighborhood of the point.

DEFINITION 1.2 *Let E and E' be two complex Banach spaces, and let Δ be an open set in the complex plane. For each λ in Δ let A_λ be a bounded linear operation with domain E and counterdomain contained in E', and let $A_\lambda x$, for each element $x \in E$, define an analytic function of λ in Δ (the values of the function lying in E'). We shall then call A_λ a linear operation analytic in Δ.*

Such operations occur frequently. Perhaps the most interesting special cases are those in which $E = E'$, and E' is the complex number system, respec-

---

* National Research Fellow.

[1] See for instance F. Riesz, *Les systèmes d'équations linéaires à une infinité d'inconnues*, Paris, 1913, Chapter IV, especially pp. 105–120; also M. H. Stone, *Linear Transformations in Hilbert Space*, New York, 1932, pp. 139–141 and p. 149.

[2] N. Wiener, Fundamenta Mathematicae, vol. 4 (1923) pp. 36–143.

[3] See the references in the author's paper, Annali di Pisa, vol. VI (1937) pp. 277–292.

tively. In the latter case we shall call A_λ a linear functional. For our purposes the most important consequence of Definition 1.2 is the following *uniformity* property: if A_λ is analytic in Δ, and for $x \in E$, $\lambda \in \Delta$

$$\lim_{\lambda' \to \lambda} \frac{A_{\lambda'} x - A_\lambda x}{\lambda' - \lambda} = A_\lambda^{(1)} x \tag{1.1}$$

then $A_\lambda^{(1)}$ is, by a well known theorem of Banach, again a bounded linear operation; *we assert that the convergence in* (1.1) *is uniform for all* x *such that* $\| x \| \leqq 1$. This means that A_λ defines a function whose values lie in a certain space of operations, and that this function is analytic in Δ (Theorem 2.1). In particular, if A_λ is a linear functional, A_λ defines an analytic function on Δ to the "conjugate space" of E (see §2).

We first examine analytic linear functionals in some of the well known Banach spaces. This also furnishes information as to the nature of analytic functions whose values lie in one of these spaces. In §4 we consider interspace transformations A_λ. For instance, if A_λ is a linear transformation in Hilbert space, defined by a bounded infinite matrix $\{a_{ik}(\lambda)\}$, it is analytic in Δ if and only if each $a_{ik}(\lambda)$ is analytic in Δ and the matrix is *uniformly bounded* on each bounded, closed subset of Δ. This conforms to the situations which arise in practice.[4]

There is also included in §4 a treatment of homogeneous and non-homogeneous linear, first order differential equations in an unknown function $x(\lambda)$ whose values lie in a Banach space E. This treatment includes many of the known results on systems of differential equations in an infinite number of unknowns.

Finally, in §5 there are presented briefly some of the most important properties of the inverse operation $(I - A_\lambda)^{-1}$ (in case it exists), together with a discussion of the resolvent when A_λ is a transformation of finite norm in Hilbert space. This last is not really new, for the theory is essentially contained in a paper by Tamarkin (see footnote (17), §5). However, the point of view is more general, and the discussion is more transparent in the abstract form.

2. The Basic Theorem

We first indicate our notations. Δ denotes some given open set of the complex plane. E, E' denote complex Banach spaces. $[E, E']$ denotes the Banach space of all linear operations A whose domain is E and whose counterdomain lies in E', the norm being given by

$$\| A \| = \text{l.u.b.}_{x \neq 0} \frac{\| Ax \|}{\| x \|} \qquad x \in E. \tag{2.1}$$

We shall always understand *linear* to imply the continuity (boundedness) of the operation. If in particular E' is the complex number system we write E^* for $[E, E']$ and call E^* the conjugate space of E. We use the usual notations l_p, (m), (c), L_p, M, C for the more familiar spaces of infinite sequences, or func-

[4] See for instance A. Wintner, Mahematische Zeitschrift, vol. 30 (1929) pp. 237–248.

tions defined on the interval (0, 1), respectively, the members of the sequences and the values of the functions being in general complex, however. For the norms in these spaces see Banach, *Opérations Linéaires*, pp. 10–12.

The basic theorem of the paper is the following.

THEOREM 2.1. *For each λ in Δ let A_λ be a member of $[E, E']$, and let A_λ be analytic in Δ in the sense of Def.* 1.2. *Then A_λ, as a function on Δ to $[E, E']$, is analytic in Δ. The converse is also true.*

PROOF: Since $\| A \| = \text{l.u.b.} \| Ax \|$ for $\| x \| \leqq 1$, the theorem may be restated as follows: if the limit in (1.1) exists for each x in E and each λ in Δ, then the convergence of the difference quotient to its limit is uniform in the unit sphere $\| x \| \leqq 1$. $A_\lambda^{(1)}$ is also linear and analytic in Δ, and the same is true of $A_\lambda^{(n)} x = \dfrac{d^n}{d\lambda^n} A_\lambda x$. Now let λ_0 be any point of Δ, and $|\lambda - \lambda_0| \leqq r$ a closed circular region which with its boundary C lies in Δ. Then

$$A_{\lambda_0}^{(n)} x = \frac{n!}{2\pi i} \int_C \frac{A_t x}{(t - \lambda_0)^{n+1}}\, dt.$$

(For the use of Cauchy's integral formula in such cases see Wiener, loc. cit. (2). Now by a theorem of Kerner[5] we can find positive constants $r_0 (r_0 \leqq r)$, M_0 such that $\| A_\lambda \| \leqq M_0$ when $|\lambda - \lambda_0| \leqq r_0$. If C_0 denotes the circle $|\lambda - \lambda_0| = r_0$ we have

$$\| A_{\lambda_0}^{(n)} x \| = \left\| \frac{n!}{2\pi i} \int_{C_0} \frac{A_t x}{(t - \lambda_0)^{n+1}}\, dt \right\| \leqq \frac{n!\, M_0 \| x \|}{r_0^n}. \tag{2.2}$$

But, x being fixed, the Cauchy-Taylor expansion

$$A_\lambda x = A_{\lambda_0} x + \sum_1^\infty \frac{(\lambda - \lambda_0)^n}{n!} A_{\lambda_0}^{(n)} x \tag{2.3}$$

is valid when $|\lambda - \lambda_0| \leqq r_0$. Moreover, on account of (2.2) the convergence is uniform in both λ and x in the sets $|\lambda - \lambda_0| \leqq \theta r_0$, $\| x \| \leqq 1$ $(0 < \theta < 1)$. From this the result easily follows. The converse is obvious.

As a consequence of Theorem 2.1 it is possible to write the equation

$$A_\lambda = A_{\lambda_0} + \sum_1^\infty \frac{(\lambda - \lambda_0)^n}{n!} A_{\lambda_0}^{(n)}. \tag{2.4}$$

Its meaning should be carefully distinguished from that of (2.3). The limit implied in (2.4) is in terms of the metric in the space $[E, E']$, whereas the convergence of (2.3) is according to the metric of E'.

Since we know now that A_λ is continuous as a function on Δ to $[E, E']$ we know that $\| A_\lambda \|$ is bounded on any bounded, closed subset of Δ. It is then

[5] M. Kerner, Studia Mathematica, vol. 3 (1931) p. 159. It results from this that $A_\lambda x$ is continuous in both variables λ, x together.

easy to see that in (2.2) we may understand r_0 to be any radius such that the closed circular region $|\lambda - \lambda_0| \leqq r_0$ lies in Δ. Hence (2.4) has the same radius of convergence as (2.3). It is also justifiable to write $A_\lambda^{(1)} = \dfrac{d}{d\lambda} A_\lambda$.

As an example of the content of the theorem consider the linear functional in the Hilbert space l_2 :

$$A_\lambda x = \sum_{n=1}^{\infty} a_n(\lambda) x_n .$$

By definition it is analytic if and only if for each sequence $\{x_n\}$ such that $\sum |x_n|^2 < \infty$ the series on the right defines an analytic function of λ in Δ. Evidently each $a_n(\lambda)$ must be analytic in Δ. The theorem asserts that for each point λ_0 in Δ

$$\lim_{\lambda \to \lambda_0} \sum_{n=1}^{\infty} \left| \frac{a_n(\lambda) - a_n(\lambda_0)}{\lambda - \lambda_0} - a_n'(\lambda_0) \right|^2 = 0.$$

The precise conditions which must be imposed on the $a_n(\lambda)$ are set forth in §3.

It is sometimes useful to write the inequality (2.2) in the form

$$\left\| \frac{1}{n!} A_{\lambda_0}^{(n)} \right\| \leqq M_{(\lambda_0)}^{n+1} \tag{2.5}$$

by taking $M(\lambda_0) = \max (M_0 , 1/r_0)$.

Given a linear operation A_λ on E to E', analytic in Δ, we can form the so-called conjugate operation[6] A_λ^* on E'^* to E^*. It will also be analytic in Δ. Hence the same is true of its conjugate A_λ^{**}, defined on E^{**} to E'^{**}. This may be regarded as an extension of the transformation A_λ, for E is isomorphic with a subset of E^{**} in such a way that if x and ξ are corresponding elements, $x \in E$, $\xi \in E^{**}$, then $A_\lambda x$ and $A_\lambda^{**} \xi$ correspond in a similar isomorphism between E' and E'^{**}. (Two spaces are isomorphic if there exists a one to one transformation, linear and continuous both ways, between them.) Finally we observe that if A_λ is analytic $\dfrac{dA_\lambda^*}{d\lambda} = \left(\dfrac{dA_\lambda}{d\lambda}\right)^*$. These facts may be verified without difficulty on the basis of Theorem 2.1.

3. Linear Functionals

In this section we shall consider linear functionals X_λ and undertake to determine, in certain special cases, the conditions under which a functional will be analytic in a given region Δ. Theorem 2.1 affords an alternative point of view which will be of great use to us. For if E is a given space its conjugate space E^* is frequently isomorphic with, or perhaps even equivalent to, a well known

[6] See Banach, *Opérations Linéaires*, Warsaw, 1932, pp. 99–100. The relation $\| A \| = \| A^* \|$ holds. This is because theorems 2 and 3 in Banach, p. 55, remain true if the spaces and functionals involved are all complex linear. See H. F. Bohnenblust and A. Sobczyk, Bulletin of the American Mathematical Society, vol. 44 (1938) p. 91–93.

Banach space whose properties we can utilize. (Two spaces E, E_1 are said to be equivalent ($E \cong E_1$) if there exists a biunique linear transformation taking E into E_1 in such a way as to preserve the norm. See Banach, loc. cit. p. 180.) If X_λ is a linear functional on a space E, and $\mathfrak{E}$ is a Banach space isomorphic with E^*, the linear correspondence between E^* and $\mathfrak{E}$ defines a function $\xi(\lambda)$ on Δ to $\mathfrak{E}$, with $\xi(\lambda)$ corresponding to X_λ, and

$$m \,||\, \xi(\lambda) \,|| \leqq ||\, X_\lambda \,|| \leqq M \,||\, \xi(\lambda) \,||, \tag{3.1}$$

where m, M, are positive constants, which are both equal to unity if the isomorphism is an equivalence. X_λ will be analytic in Δ if and only if $\xi(\lambda)$ is analytic in Δ.

The following equivalences are well known:[7] $l_p \cong (l_{p/p-1})^*$ $(p > 1)$; $l \cong (c)^*$; $(m) \cong l^*$; $L_p \cong L_{p/p-1}$ $(p > 1)$; $M \cong L^*$. Most of our results will be stated as theorems about functions on Δ to $\mathfrak{E}$. Their interpretation in terms of linear functionals will be evident from the proofs.

Our problem is easily solved for the sequence spaces. The results are collected in the next theorem.

THEOREM 3.1. *To a function $\xi(\lambda)$ on Δ to any one of the spaces $l_p (p \geqq 1)$, (m), (c) there corresponds a sequence $\{\xi_n(\lambda)\}$ of complex-valued functions. In order that $\xi(\lambda)$ be analytic in Δ it is first of all necessary that each $\xi_n(\lambda)$ be analytic in Δ, and secondly, in the several cases it is necessary that*

l_p : (i) *the series $\sum_n |\, \xi_n(\lambda) \,|^p$ be convergent at each point of Δ, and bounded in each closed, bounded sub-region of Δ,*[8]

(m): (ii) $|\, \xi_n(\lambda) \,| < K(\Delta_0)$ *in* Δ_0 $(n = 1, 2, \cdots)$,

(c) : (iii) *condition* (ii) *hold, and* $\lim \xi_n(\lambda)$ *as* $n \to \infty$ *exist at each point of* Δ. *Conversely, if each $\xi_n(\lambda)$ is analytic in Δ the sequence defines an analytic function on Δ to one of the spaces l_p, (m), (c) if the appropiate one of the above three conditions holds.*

PROOF: As in ordinary analysis $||\, \xi(\lambda) \,|| < K(\Delta_0)$ on Δ_0 if $\xi(\lambda)$ is analytic (and hence continuous) in Δ, with values in a Banach space. Moreover, for each of the above sequence spaces we have the inequality $|\, \xi_n \,| \leqq ||\, \xi \,||$. From these remarks the necessity of the conditions follows. As to their sufficiency we note that according as $\mathfrak{E}$ is $l_p (p > 1)$, l, or (m), the linear functional corresponding to $\xi(\lambda)$ is of the form

$$X_\lambda(x) = \sum_{n=1}^{\infty} \xi_n(\lambda) x_n \qquad x \in l_{p/p-1}, \quad \xi(\lambda) \in l_p \quad (p > 1)$$

$$X_\lambda(x) = \sum_{n=0}^{\infty} \xi_{n+1}(\lambda) x_n \qquad x \in (c), \quad \lim_{n \to \infty} x_n = x_0, \quad \xi(\lambda) \in l$$

$$X_\lambda(x) = \sum_{n=1}^{\infty} \xi_n(\lambda) x_n \qquad x \in l, \quad \xi(\lambda) \in (m).$$

[7] For the corresponding *real* spaces these facts are all contained in Banach, loc. cit. Chapter IV. The forms of the general linear functionals are the same for the complex spaces.

[8] For convenience and brevity the statement that a numerical function $F(\lambda)$ is bounded

Consider first $\xi(\lambda) \in l$. If $\xi_n(\lambda)$ is analytic in Δ and condition (i) is satisfied ($p = 1$), the functional $X_\lambda(x)$ is analytic in Δ for each $x \in (c)$. In fact, the series converges, and

$$\left| \sum_{n=0}^{N} \xi_{n+1}(\lambda) x_n \right| \leqq \max_{n \geqq 1} |x_n| \sum_{n=0}^{N} |\xi_{n+1}(\lambda)| < \| x \| K(\Delta_0)$$

in Δ_0. Therefore, by a well known theorem of Montel, $X_\lambda(x)$ is analytic. The demonstration for $l_p (p > 1)$ or (m) is similar, and will be omitted. The space (c) is a subspace of (m) (with the same norm). Hence for $\xi(\lambda)$ on Δ to (c) to be analytic in Δ it is sufficient to know that it is analytic as a function on Δ to (m), and that the values of both $\xi(\lambda)$ and its derivative $\xi'(\lambda)$ lie in (c). The conditions (ii) and (iii) imply that this is so when $\xi_n(\lambda)$ is analytic in Δ, for then $\lim \xi_n'(\lambda)$ as $n \to \infty$ exists at each point in Δ. In all cases the derivative $\xi'(\lambda)$ is defined by the sequence $\{\xi_n'(\lambda)\}$.

We turn now to the more interesting and difficult theorems corresponding to Theorem 3.1 for the usual function spaces. Analytic functions on Δ to L_2 were considered by Wiener (see footnote 2). He outlined a proof of Lemma 1 (below), but did not go as far as our Theorem 3.2. What Bôcher[9] calls semi-analytic functions are essentially analytic functions on Δ to C. We begin with functions on Δ to L_p. We adopt the view that the Banach space L_p has for its elements the classes of equivalent functions ξ, where ξ is measureable, and $|\xi|^p$ integrable, two functions being equivalent if they are almost everywhere equal on the interval I: $0 \leqq s \leqq 1$. A function $F(\lambda)$ on Δ to L_p then orders to each λ a *class* of functions. A particular member of this class may be denoted by $\xi_\lambda(s)$. We shall call it a *representative* of $F(\lambda)$. It is defined except possibly on a set of measure zero depending on λ. If for each λ a representative ξ_λ is selected, the aggregate $\xi_\lambda(s)$, as λ varies, may not define a function of the two variables s, λ in the usual sense. It is conceivable that there may not exist any s for which *all* the $\xi_\lambda(s)$ are defined. What we shall show is that if $F(\lambda)$ is analytic in Δ it is possible to make the above selection in such a way as to arrive at a function $\xi(s, \lambda)$ defined for each λ in Δ and each s not in a certain *fixed* set of measure zero. This function is analytic in Δ for each such s, and for each λ it is a representative of $F(\lambda)$. The precise statement is contained in Theorem 3.2, along with a converse.

We begin by proving a lemma.

LEMMA 1. *If $F(\lambda)$ is an analytic function on Δ to $L_p (p \geqq 1)$, then to each point λ_0 in Δ corresponds a positive number $\delta(\lambda_0)$, a set $E(\lambda_0) \subset I$ of measure zero, and a function $\bar{\xi}(s, \lambda)$ with the properties*

1° *$\bar{\xi}(s, \lambda)$ is defined when $|\lambda - \lambda_0| < \delta(\lambda_0)$ and $s \in I - E(\lambda_0)$*
2° *for each $s \in I - E(\lambda_0)$ $\bar{\xi}(s, \lambda)$ is analytic, $|\lambda - \lambda_0| < \delta(\lambda_0)$*
3° *for each λ, $|\lambda - \lambda_0| < \delta(\lambda_0)$, $\bar{\xi}(s, \lambda)$ is a representative of $F(\lambda)$.*

---

in each closed, bounded sub-region of Δ will be expressed symbolically by writing: $|F(\lambda)| < K(\Delta_0)$ in Δ_0. Here Δ_0 denotes an arbitrary closed, bounded sub-region of Δ, and $K(\Delta_0)$ some positive constant depending on Δ_0, but not on λ.

[9] M. Bôcher, Annals of Mathematics, vol. 12 (1911) pp. 18-26.

PROOF: Let λ_0 be a point in Δ. Since $F(\lambda)$ is analytic we may write, for some positive ρ

$$F(\lambda) = \sum_{n=0}^{\infty} (\lambda - \lambda_0)^n \frac{F^{(n)}(\lambda_0)}{n!} \qquad |\lambda - \lambda_0| < \rho. \tag{3.2}$$

This means, if $\xi_n(s)$ is any one of the representatives of the point $\frac{1}{n!} F^{(n)}(\lambda_0)$ in L_p, and $\xi_\lambda(s)$ any representative of $F(\lambda)$, that

$$\lim_{n\to\infty} \int_0^1 \left| \xi_\lambda(s) - \sum_{\nu=0}^{n} (\lambda - \lambda_0)^\nu \xi_\nu(s) \right|^p ds = 0, \qquad |\lambda - \lambda_0| < \rho. \tag{3.3}$$

From (3.2) it follows that if $0 < \rho_1 < \rho$ there exists a constant B such that

$$\left\| \rho_1^n \frac{1}{n!} F^{(n)}(\lambda_0) \right\|^p = \int_0^1 |\rho_1^n \xi_n(s)|^p ds < B \qquad n = 0, 1, \cdots \tag{3.4}$$

Now let θ, A be positive constants, and $0 < \theta < 1$. Let $E_n(A)$ be the set of points on the interval I for which ξ_n is undefined, or $|\rho_1^n \xi_n(s)| \geqq \theta^{-n/2} A$. The measure of $E_n(A)$ cannot exceed $BA^{-p}\theta^{np/2}$, because of (3.4). Hence if $E(A) = \sum_n E_n(A)$ we conclude that

$$mE(A) \leqq \frac{B}{A^p(1 - \theta^{p/2})}. \tag{3.5}$$

Thus on the set $I - E(A)$ $|\rho_1^n \xi_n(s)| < \theta^{-n/2} A$, whence it follows that if $|\lambda - \lambda_0| \leqq \theta\rho_1$

$$\left| \sum_{\nu=n}^{n+k} (\lambda - \lambda_0)^\nu \xi_\nu(s) \right| \leqq \sum_{\nu=n}^{n+k} \theta^{\nu/2} A < \frac{\theta^{n/2} A}{1 - \theta^{\frac{1}{2}}}.$$

Therefore the series

$$\sum_{0}^{\infty} (\lambda - \lambda_0)^n \xi_n(s) \tag{3.6}$$

converges absolutely and uniformly for all pairs (s, λ) such that $s \in I - E(A)$, $|\lambda - \lambda_0| \leqq \theta\rho_1$. In the same way we see that for these pairs the partial sums of the series (3.6) are bounded.

Now let A take on the values $N = 1, 2, \cdots$. Evidently $E_n(N+1) \subseteq E_n(N)$ and $E(N+1) \subseteq E(N)$. Also $\lim mE(N) = 0$ as $N \to \infty$. Consider the set $E = \prod_{N=1}^{\infty} E(N)$. Clearly $mE = 0$, and if $s \in I - E$ then $s \in I - E(N)$ for all sufficiently large N. From this we see that the series (3.6) defines a function $\bar{\xi}(s, \lambda)$ satisfying conditions 1° and 2° of the lemma, if we take $\delta(\lambda_0) = \theta\rho_1$, $E(\lambda_0) = E$. It also satisfies condition 3° as we shall show. Let $|\lambda - \lambda_0| < \theta\rho_1$. Then the partial sums of the series (3.6) converge boundedly to $\bar{\xi}(s, \lambda)$ on $I - E(N)$, and hence

$$\lim_{n\to\infty} \int_{I-E(N)} \left| \bar{\xi}(s, \lambda) - \sum_{\nu=0}^{n} (\lambda - \lambda_0)^\nu \xi_\nu(s) \right|^p ds = 0.$$

By (3.3) we know therefore that $\xi_\lambda(s) = \bar{\xi}(s, \lambda)$ at least in a set $\{I - E(N)\} - e_N(\lambda)$, where $e_N(\lambda)$ is a set of measure zero contained in $I - E(N)$. Hence if we define $e(\lambda) = \sum_N e_N(\lambda)$, $e(\lambda) + E$ is of measure zero, and $\xi_\lambda(s) = \bar{\xi}(s, \lambda)$ when $s \in I - (E + e(\lambda))$. This completes the proof of the lemma.

We can now formulate the analogue of Theorem 3.1.

THEOREM 3.2. *If $F(\lambda)$ is an analytic function on Δ to $L_p (p \geqq 1)$ there exists a function $\bar{\xi}(s, \lambda)$ with the properties:*

1° *$\bar{\xi}(s, \lambda)$ is defined when λ is in Δ and $s \in I - E$, where E is a set of measure zero independent of λ;*

2° *for each $s \in I - E$ $\bar{\xi}(s, \lambda)$ is analytic in Δ;*

3° *for each λ in Δ $\bar{\xi}(s, \lambda)$ is in L_p, and*

$$\int_0^1 | \bar{\xi}(s, \lambda) |^p \, ds < K(\Delta_0) \text{ in } \Delta_0 \,^{8}$$

4° *$\bar{\xi}(s, \lambda)$ is, for each λ, a representative of $F(\lambda)$.*

Conversely, if $\bar{\xi}(s, \lambda)$ is a function satisfying 1° – 3°, and a function $F(\lambda)$ on Δ to L_p is defined by requiring 4° to hold, then $F(\lambda)$ is analytic in Δ.

PROOF: The first part of the theorem follows readily from Lemma 1. We have, corresponding to each point μ in Δ, a set $E(\mu)$ of measure zero, a circular region K_μ: $| \lambda - \mu | < \delta(\mu)$, and a function $\bar{\xi}_\mu(s, \lambda)$ with the properties described in the lemma. We can select a sequence $\{\mu_n\}$ of points in Δ such that the circles $\{K_{\mu_n}\}$ completely cover Δ. Let $E' = \sum E(\mu_n)$; E' is of measure zero. Consider now two of the circles, say K_{μ_m}, K_{μ_n}, which overlap. Let λ be a rational point of the intersection $K_{\mu_m} \cdot K_{\mu_n}$. Since $\bar{\xi}_{\mu_m}(s, \lambda)$ and $\bar{\xi}_{\mu_n}(s, \lambda)$ are both representatives of $F(\lambda)$ they are equal except possibly on a set of measure zero (depending on λ). Enumerating the rational points of $K_{\mu_m} \cdot K_{\mu_n}$ it follows that $\bar{\xi}_{\mu_m}(s, \lambda) = \bar{\xi}_{\mu_n}(s, \lambda)$ for all these rational points, except possibly on a set of measure zero depending only on m and n, say e_{mn}. Since for fixed s the functions are analytic in $K_{\mu_m} \cdot K_{\mu_n}$ the above equality remains true for *all* points of $K_{\mu_m} \cdot K_{\mu_n}$. Finally, discarding the null set $\sum_{m,n} e_{m,n}$ (which includes E') we conclude that except for a set of measure zero, *independent of* m, n, λ, $\bar{\xi}_{\mu_m}(s, \lambda) = \bar{\xi}_{\mu_n}(s, \lambda)$ whenever $\lambda \in K_{\mu_m} \cdot K_{\mu_n}$. It is then easy to see that the functions $\{\bar{\xi}_{\mu_n}(s, \lambda)\}$ determine a function $\bar{\xi}(s, \lambda)$ of the required specifications 1° – 4°. 3° is a consequence of 4°, together with the fact that, $F(\lambda)$ being analytic in Δ, $|| F(\lambda) || < K(\Delta_0)$ in Δ_0.

To prove the converse we shall suppose first that $p > 1$. In view of the remarks at the beginning of this section $F(\lambda)$ will be analytic if the linear functional

$$X_\lambda(x) = \int_0^1 \bar{\xi}(s, \lambda) x(s) \, ds \qquad x \in L_{p/p-1} \tag{3.7}$$

is analytic in Δ, since $\bar{\xi}(s, \lambda)$ is a representative of $F(\lambda)$. The proof that $X_\lambda(x)$ is analytic will be furnished by the following lemma, the proof of which will also be of use in treating the case $p = 1$.

LEMMA 2. *Let $f(s, \lambda)$ be a function with the properties 1° – 3° of Theorem 3.2 for $p = 1$. Then $\dfrac{\partial f(s, \lambda)}{\partial \lambda}$ is in L for each λ; the integral $\displaystyle\int_0^1 f(s, \lambda)\, ds$ defines an analytic function in Δ, and*

$$\frac{d}{d\lambda}\int_0^1 f(s, \lambda)\, ds = \int_0^1 \frac{\partial f(s, \lambda)}{\partial \lambda}\, ds. \tag{3.8}$$

PROOF. Let λ_0 be in Δ. Then for some positive ρ independent of s, and $s \in I - E$

$$f(s, \lambda) = \sum_{n=0}^{\infty} (\lambda - \lambda_0)^n f_n(s) \qquad |\lambda - \lambda_0| < \rho$$

$$f_n(s) = \frac{1}{n!}\left[\frac{\partial^n f(s, \lambda)}{\partial \lambda^n}\right]_{\lambda=\lambda_0}. \tag{3.9}$$

It is clear that each $f_n(s)$ is measurable. To prove that they are in L, let $0 < \rho_1 < \rho$, and write $\varphi(s, \theta) = f(s, \lambda_0 + \rho_1 e^{i\theta})$ $(0 \leqq \theta \leqq 2\pi)$. Then from (3.9) we obtain

$$\varphi(s, \theta) = \sum_{0}^{\infty} \rho_1^n e^{in\theta} f_n(s) \qquad s \in I - E. \tag{3.10}$$

Since the terms of this series are measurable in the two-dimensional set $0 \leqq s \leqq 1$, $0 \leqq \theta \leqq 2\pi$, the same is true of $\varphi(s, \theta)$. It is further true that $\varphi(s, \theta)$ is *integrable* over this set. For, since $\displaystyle\int_0^1 |\varphi(s, \theta)|\, ds$ exists and is bounded for $0 \leqq \theta \leqq 2\pi$ it defines an integrable function of θ. Hence $|\varphi(s, \theta)|$, and also $\varphi(s, \theta)$ is integrable.[10] Moreover, since

$$f_n(s) = \frac{1}{2\pi}\int_0^{2\pi} \frac{\varphi(s, \theta)}{\rho_1^n e^{in\theta}}\, d\theta \tag{3.11}$$

we conclude, by the theorem of Fubini, that $f_n(s)$ is integrable. Now consider the expression

$$\int_0^1 \left[\frac{f(s, \lambda) - f(s, \lambda_0)}{\lambda - \lambda_0} - f_1(s)\right] ds = \frac{1}{2\pi}\int_0^1 ds \int_0^{2\pi} \frac{(\lambda - \lambda_0)\varphi(s, \theta)\, d\theta}{(t - \lambda)(t - \lambda_0)}$$

$$t - \lambda_0 = \rho_1 e^{i\theta}$$

$$|\lambda - \lambda_0| = \rho_1/2$$

obtained by use of Cauchy's formula. Since $\varphi(s, \theta)$ is integrable we may invert the order of integration, thus getting an estimate for the expression on the right:

$$\left|\frac{\lambda - \lambda_0}{2\pi}\int_0^{2\pi} \frac{d\theta}{(t - \lambda)(t - \lambda_0)} \int_0^1 \varphi(s, \theta)\, ds\right| \leqq \frac{|\lambda - \lambda_0|}{\frac{1}{2}\rho_1^2} \operatorname*{l.u.b.}_{0 \leqq \theta \leqq 2\pi} \int_0^1 |\varphi(s, \theta)|\, ds.$$

[10] C. Carathéodory, *Vorlesungen über Reelle Funktionen*, Leipzig, 1927, p. 637.

The quantity on the right is finite, and tends to zero with $|\lambda - \lambda_0|$. Since λ_0 was arbitrary (3.8) has been established, and the lemma proved.

The converse portion of Theorem 3.2 is now proved when $p > 1$, for we may apply Lemma 2 to our consideration of the integral (3.7), with the aid of Hölder's inequality. It remains to give a proof when $p = 1$. We shall do this directly, since the previous method is not available. Now $\bar{\xi}(s, \lambda)$ satisfies the hypotheses of Lemma 2. Let $\xi_\lambda(s)$ be one of the representatives of $F(\lambda)$, and λ_0 a point in Δ. As in the proof of Lemma 2 we may write

$$\left.\begin{aligned} \bar{\xi}(s, \lambda) &= \sum_0^\infty (\lambda - \lambda_0)^n \bar{\xi}_n(s) && |\lambda - \lambda_0| < \rho \\ \bar{\xi}_n(s) &= \frac{1}{2\pi} \int_0^{2\pi} \frac{\varphi(s, \theta)}{\rho_1^n e^{in\theta}}\, d\theta && s \epsilon I - E \end{aligned}\right\}$$

where $\varphi(s, \theta) = \bar{\xi}(s, \lambda_0 + \rho_1 e^{i\theta})$ $(\rho_1 < \rho)$. We shall prove that

$$\lim_{n \to \infty} \int_0^1 \left| \xi_\lambda(s) - \sum_{\nu=0}^{n} (\lambda - \lambda_0)^\nu \bar{\xi}_\nu(s) \right| ds = 0 \tag{3.12}$$

for all sufficiently small values of $|\lambda - \lambda_0|$. We may replace $\xi_\lambda(s)$ by $\bar{\xi}(s, \lambda)$ in the left member of (3.12). Then instead of proving (3.12) it will be sufficient to prove that

$$\lim_{n \to \infty} \sum_{\nu=n}^{\infty} |\lambda - \lambda_0|^\nu \int_0^1 |\bar{\xi}_\nu(s)|\, ds = 0. \tag{3.13}$$

But

$$\int_0^1 |\bar{\xi}_\nu(s)|\, ds \leqq \frac{1}{2\pi} \int_0^1 ds \int_0^{2\pi} \frac{|\varphi(s, \theta)|}{\rho_1^\nu}\, d\theta \leqq \frac{K}{\rho_1^\nu}$$

where K is some positive constant, because of condition 3° and the legitimacy of changing the order of integration. Hence (3.13) and (3.12) are true if $|\lambda - \lambda_0| < \rho_1$. But (3.12) expresses precisely the fact that $F(\lambda)$ is analytic at the point λ_0, which was arbitrary. The proof is thus complete.

There is a theorem very similar to Theorem 3.2 for analytic functions on Δ to M, the space of measurable, almost—bounded functions. We pass over it, however, to the more interesting problem of linear functionals defined on the space C of continuous functions. As is well known, such a functional is expressible as a Stieltjes integral with the aid of a function of bounded variation. For our purposes we lay down the following definition.

DEFINITION. *Denote by BV the class of all complex valued functions ξ defined on* (0, 1) *and of finite total variation $V(\xi)$ over this interval. If $\xi \epsilon BV$ define a new function*

$$\xi_*(0) = 0 \qquad \xi_*(1) = \xi(1) - \xi(0)$$

$$\xi_*(s) = \xi(s + 0) - \xi(0) \qquad 0 < s < 1.$$

A function $\xi \in BV$ is said to be in normal form if it is identical with ξ_*. The class of all members of BV which are in normal form will be denoted by $(BV)_*$.

It is readily proved that ξ is in normal form if and only if $\xi(0) = 0$ and $\xi(s + 0) = \xi(s)$ $(0 < s < 1)$. BV is a Banach space when it is normed by

$$\| \xi \| = | \xi(0) | + V(\xi). \tag{3.14}$$

$(BV)_*$ is also a Banach space with this same norm: $\| \xi_* \| = V(\xi_*)$. Observe that $| \xi(s) | \leqq \| \xi \|$.

It is not difficult to prove that $(BV)_*$ is equivalent to the conjugate space C^* and hence that if $\xi_* \in (BV)_*$ corresponds to $X \in C^*$ then

$$\| X \| = \| \xi_* \|. \tag{3.15}$$

We shall not prove this equivalence here. By means of it we may regard a linear functional X_λ on C as defined by a function $\xi_*(s, \lambda) \in (BV)_*$ for each λ. X_λ will be analytic in Δ if and only if ξ_* defines a function analytic in the metric of $(BV)_*$.

THEOREM 3.3. *The linear functional on the space C,*

$$X_\lambda(x) = \int_0^1 x(s)\, d_s \xi_*(s, \lambda) \tag{3.16}$$

where $\xi_*(s, \lambda) \in (BV)_*$ for each λ in Δ, is analytic in Δ if and only if

1° $\xi_*(s, \lambda)$ is analytic in Δ for each s, $0 \leqq s \leqq 1$,

2° $V(\xi_*) < K(\Delta_0)$ in Δ_0.[8]

Under these circumstances the set of points σ on $(0, 1)$ such that $\xi_*(s, \lambda)$ is discontinuous at σ for some λ in Δ is enumerable.

PROOF: The necessity of 1°, 2° is quite evident from the relations $| \xi_*(s, \lambda) | \leqq V(\xi_*) = \| \xi_* \|$ and the remarks preceding the theorem. To prove the converse let $x(s)$ be any continuous function and choose a sequence of subdivisions of the interval $(0, 1)$, say by points $0 = s_0^{(n)} < s_1^{(n)} < \cdots < s_n^{(n)} = 1$, in such a manner that for each λ in Δ

$$\int_0^1 x(s)\, d_s \xi_*(s, \lambda) = \lim_{n \to \infty} \sum_{i=1}^{n} x(s_i^{(n)})[\xi_*(s_i^{(n)}, \lambda) - \xi_*(s_{i-1}^{(n)}, \lambda)]. \tag{3.17}$$

By 2° we see that the sums on the right are uniformly bounded in Δ_0. Since the individual terms are analytic in Δ, by 1°, it follows that the limit function is analytic in Δ. To prove the concluding assertion we make use of the series expansion,

$$\xi_*(s, \lambda) = \sum_0^\infty \frac{(\lambda - \lambda_0)^n}{n!} \left\{ \frac{\partial^n \xi_*(s, \lambda)}{\partial \lambda^n} \right\} \lambda = \lambda_0 \tag{3.18}$$

which converges in the metric of $(BV)_*$, and hence uniformly with respect to s. Therefore, for each λ in a certain circle with center at λ_0 the discontinuities of $\xi_*(s, \lambda)$ can be at most those of the terms on the right. But these terms are in $(BV)_*$, and their discontinuities, which are at most enumerable, are independent of λ. If we cover Δ with an enumerable set of such circles the result easily follows.

It is of some interest to consider conditions on a function $\xi(s, \lambda)$ which is in BV for each $\lambda \in \Delta$, sufficient to insure that its normal form $\xi_*(s, \lambda)$ have the properties listed in Theorem 3.3. It is clearly sufficient to require that

(i) $\xi(0, \lambda)$ and $\xi(1, \lambda)$ be analytic in Δ,

(ii) there exist a set S everywhere dense on $(0, 1)$, and independent of λ such that for each point $s \in S$, $\xi(s, \lambda)$ is analytic in Δ,

(iii) the total variation of $\xi(s, \lambda)$, calculated for subdivisions of the interval whose end-points lie in S, is bounded in regions Δ_0.

4. Linear Transformations

The results of §3 may be used to advantage in studying linear operations other than functionals. For instance, if E' is one of the sequence spaces l_p, (m) or (c), an operation A_λ on E to E' is defined by a sequence $\{a_\lambda^{(i)}\}$ of linear functionals on E. Because of the relations $| a_\lambda^{(i)} x | \leqq || A_\lambda x ||$ valid for these choices of E' we have $|| a_\lambda^{(i)} || \leqq || A_\lambda ||$. It follows readily that A_λ will be analytic if and only if each $a_\lambda^{(i)}$ is an analytic linear functional, and $|| A_\lambda || < K(\Delta_0)$ in Δ_0.[8] In case $E = E' = l_2$ (Hilbert space) this requirement is, in view of Theorem 3.1, that if A_λ is defined by the matrix $\{a_{ik}(\lambda)\}$ then $a_{ik}(\lambda)$ must be analytic in Δ, and the matrix *uniformly* bounded (its bound is $|| A_\lambda ||$) in regions of the type Δ_0. It is to be observed that a singular point of $A_\lambda = \{a_{ik}(\lambda)\}$ can occur even though all the $a_{ik}(\lambda)$ individually are regular at that point.[11] The same is of course true of a linear functional $X_\lambda(x) = \sum a_n(\lambda) x_n$ on l_2, as is shown by the example $a_n(\lambda) = \lambda^n$, the region of regularity in this case being $| \lambda | < 1$.

As an interesting application we may turn our theory to the study of differential systems of the form

$$\frac{dx}{d\lambda} = A_\lambda x \qquad x(0) = y \tag{4.1}$$

where A_λ is a linear transformation of the space E, analytic at $\lambda = 0$, and y is an arbitrarily prescribed element of E. If E is l_p this takes the form

$$\frac{dx_i}{d\lambda} = \sum_{k=1}^{\infty} a_{ik}(\lambda) x_k \qquad x_i(0) = y_i \quad i = 1, 2, \cdots \tag{4.2}$$

where the matrix $\{a_{ik}(\lambda)\}$ is subjected to certain conditions which we have already given for the case $p = 2$. Systems of this type have been considered frequently.[12] The system (4.1) also includes equations of the form

$$\frac{\partial x(s, \lambda)}{\partial \lambda} = x(s, \lambda) + \int_0^1 A(s, t, \lambda) x(t, \lambda)\, dt, \qquad x(s, 0) = y(s). \tag{4.3}$$

[11] See A. Wintner, loc. cit. p. 246. The matrix may cease to be bounded.

[12] For a number of references see R. D. Carmichael, Bulletin of the American Mathematical Society, vol. 42 (1936) pp. 211–216. For $E = l_2$ our results are essentially those of W. L. Hart, American Journal of Mathematics, vol. 39 (1917) pp. 407–424. The system (4.2) was treated by Sternberg, Heidelberg Akademie Sitzungsberichte, 1920, no. 10, pp. 1–21 under hypotheses sufficiently restrictive to make it of type (4.1) for the space $E = (m)$.

For instance, if $A(s, t, \lambda)$ is continuous for $0 \leqq s, t \leqq 1$, $|\lambda| < r$, and analytic in $|\lambda| < r$ for fixed values of (s, t), then (4.3) is a system of type (4.1) in the space C. The solution $x(s, \lambda)$ will be continuous in the range $0 \leqq s \leqq 1$, $|\lambda| < r$, and analytic in λ for each s. (See footnote (9).) If $A(s, t, \lambda)$ is a kernel of Hilbert-Schmidt type with the properties described in the remarks following Theorem 4.3, then (4.3) may be regarded as of type (4.1) in the Hilbert space L_2, and the solution will be analytic in accordance with Theorem 3.2. Still more general equations may be considered.

We turn now to the general system (4.1). Our results may be stated as follows.

THEOREM 4.1. *If A_λ is a linear transformation on E to E, analytic in the circle $|\lambda| < r$, then the system* (4.1) *admits a unique analytic solution $x(\lambda)$ in the circle $|\lambda| < r$. This solution may be written in the form $x(\lambda) = B_\lambda y$, where B_λ is a linear transformation analytic for $|\lambda| < r$.*

PROOF: We endeavor to find a formal solution by writing

$$A_\lambda = \sum_0^\infty \frac{\lambda^n}{n!} A_0^{(n)}, \qquad x(\lambda) = \sum_0^\infty \lambda^n x_n, \qquad x_0 = y \tag{4.4}$$

substituting in (4.1), and equating coefficients of like powers of λ. This leads to the equations

$$\begin{aligned} \sum_0^\infty (n+1)x_{n+1} &= \sum_{m,n=0}^\infty \frac{\lambda^{m+n}}{m!} A_0^{(m)} x_n \\ (n+1)x_{n+1} &= \sum_{K=0}^n \frac{1}{K!} A_0^{(K)} x_{n-K} \qquad n = 0, 1, \cdots. \end{aligned} \tag{4.5}$$

The rearrangement of the double series will be justified presently. We now have a means of determining the x_n successively. If we choose any $r_0 < r$ there will exist an M_0 such that (2.2) holds; we then obtain the estimates

$$(n+1)\,\|x_{n+1}\| \leqq M_0 \sum_{K=0}^n \frac{1}{r_0^K}\,\|x_{n-K}\|.$$

If we define constants $\{b_n\}$ recursively by

$$b_{n+1} = \frac{M_0}{n+1}\left[b_n + \frac{1}{r_0} b_{n-1} + \cdots + \frac{1}{r_0^{n-1}} b_1 + \frac{1}{r_0^n}\right], \qquad b_1 = M_0$$

we readily find that $\|x_{n+1}\| \leqq b_{n+1}\|x_0\|$, and

$$(n+1)b_{n+1} = b_n\left(\frac{n}{r_0} + M_0\right).$$

Hence the series $\sum b_n\lambda^n$ converges when $|\lambda| < r_0$. It follows that the formal power series defining $x(\lambda)$ in (4.4) is convergent if $|\lambda| < r$, and $x(\lambda)$ is analytic. The justification of the rearrangement in (4.4) now presents no difficulty, and we actually have a solution of (4.1). The uniqueness is proved by a familiar argument.

If we define $B_0 y = y$, $B_n y = x_n$ $(n = 1, 2, \cdots)$ we see that B_n is a linear transformation with $\| B_n \| \leqq b_n$, and from this it follows that the transformation

$$B_\lambda y = \sum_0^\infty \lambda^n B_n y \tag{4.6}$$

is analytic for $| \lambda | < r$. Since $x(\lambda) = B_\lambda y$ we see that $\frac{dB_\lambda}{d\lambda} = A_\lambda B_\lambda$, and that B_λ reduces to the identity transformation when $\lambda = 0$. If A_λ is a fixed transformation A independent of λ we see that $B_\lambda = \sum_0^\infty \frac{\lambda^n}{n!} A^n$, which we may define as $e^{\lambda A}$.

Now let us consider the non-homogeneous equation

$$\frac{dx}{d\lambda} = A_\lambda x + z(\lambda) \tag{4.7}$$

where $z(\lambda)$ is a given function with values in E, analytic in the circle $| \lambda | < r$. Concerning this equation we prove

THEOREM 4.2. *There exists a solution of* (4.7) *which is analytic for* $| \lambda | < r$ *and which vanishes at* $\lambda = 0$.

From this result and Theorem 4.1 it is clear how to obtain the general solution of (4.7). The proof depends on the consideration of the equation

$$\frac{dX}{d\lambda} = -A_\lambda^* X \tag{4.8}$$

in the conjugate space E^*, where A_λ^* is the transformation conjugate to A_λ (see §2). We recall that if X is a linear functional defined on E, $A_\lambda^* X$ is the linear functional Y defined by $Y(x) = X(A_\lambda x)$. A_λ^* is analytic, $| \lambda | < r$, and hence Theorem 4.1 applies to (4.8). Bearing this in mind we consider first a lemma.

LEMMA 3. *There exists a linear transformation* H_λ *on* E, *analytic for* $| \lambda | < r$, *and such that for each* λ *it defines a one to one linear mapping of* E *on itself, the inverse transformation* H_λ^{-1} *also being analytic,* $| \lambda | < r$, *and finally*

$$\frac{dH_\lambda}{d\lambda} = A_\lambda H_\lambda. \tag{4.9}$$

PROOF: Let H_0 be an arbitrary linear transformation on E, such that it defines a one to one mapping of E on itself. The inverse H_0^{-1} is then also linear.[13] Then the transformation $H_\lambda = B_\lambda H_0$, where B_λ is defined by (4.6), is analytic and has the property (4.9). We shall prove that it satisfies the demands of the theorem.

[13] Schauder, Studia Mathematica, vol. 2 (1930) pp. 1–6. Also Banach, loc. cit. Theorems 2 and 5, pp. 147–149.

Form the conjugate transformation $(H_0^{-1})^*$ and let $\mathcal{H}_\lambda$ be a transformation in E^*, so determined that it is analytic for $|\lambda| < r$, $\frac{d\mathcal{H}_\lambda}{d\lambda} = -A_\lambda^*\mathcal{H}_\lambda$, and $\mathcal{H}_0 = (H_0^{-1})^*$. This is possible, since Theorem 4.1 applies to (4.8). Now if Y is an arbitrary element of E^*, then $X_\lambda = \mathcal{H}_\lambda Y$ is a solution of (4.9), and therefore $\frac{d}{d\lambda} X_\lambda(x) = X_\lambda^{(1)}(x) = -X_\lambda(A_\lambda x)$ for all x in E, by the definition of A_λ^*. But if $x(\lambda)$ is a solution of (4.1) we have $x'(\lambda) = A_\lambda x(\lambda)$, and hence

$$\frac{d}{d\lambda}\{X_\lambda(x(\lambda))\} = X_\lambda^{(1)}(x(\lambda)) + X_\lambda(x'(\lambda)) = 0.^{14}$$

Hence

$$\frac{d}{d\lambda}\{X_\lambda(H_\lambda y)\} = 0 \qquad y \in E. \tag{4.10}$$

From this we infer

$$X_\lambda(H_\lambda y) = X_0(H_0 y) = Y(y) \tag{4.11}$$

since $X_0 = \mathcal{H}_0 Y = (H_0^{-1})^* Y$, whence $X_0(x) = Y(H_0^{-1}x)$. Equation (4.11) is equivalent to the equation

$$H_\lambda^* \mathcal{H}_\lambda Y = Y. \tag{4.12}$$

Now H_λ^* is analytic, and H_0^* defines a one to one mapping of E^* on itself, with inverse[15] $(H_0^{-1})^* = (H_0^*)^{-1}$. Hence, as is proved in §5 (Theorem 5.1), there exists a neighborhood of $\lambda = 0$ in which the inverse $(H_\lambda^*)^{-1}$ exists, is defined throughout E^*, and analytic. For such values of λ (4.12) implies that $\mathcal{H}_\lambda = (H_\lambda^*)^{-1}$. From this it follows that $(H_\lambda^*\mathcal{H}_\lambda - \mathcal{H}_\lambda H_\lambda^*)Y \equiv 0$ in the neighborhood of $\lambda = 0$, and so for all $|\lambda| < r$, since the left side is analytic. Thus finally we see that H_λ^* and $\mathcal{H}_\lambda$ are mutually inverse for all $|\lambda| < r$. Banach's theorem, referred to above, then assures the unique solvability of the equation $H_\lambda y = x$ for every x in E, thus defining a linear inverse $y = H_\lambda^{-1}x$. By Theorem 5.1 it is analytic in $|\lambda| < r$. This completes the proof of the lemma.

The proof of Theorem 4.2 is now a simple matter. Consider the function defined by the integral

$$w(\lambda) = \int_0^\lambda H_t^{-1} z(t)\, dt \qquad |\lambda| < r. \tag{4.13}$$

It is single-valued and analytic, since the integrand is analytic. Moreover, $w(0) = 0$ (by definition) and $w'(\lambda) = H_\lambda^{-1}z(\lambda)$. Hence if we define

$$x(\lambda) = H_\lambda w(\lambda) \tag{4.14}$$

---

[14] This differentiation formula is easily established by use of (2.3) and the expansion of $x(\lambda)$. It holds under very general conditions. See footnote (18).

[15] Banach, loc. cit. Theorem 5, p. 149. This theorem remains true in the complex case.

then $x(\lambda)$ is analytic, and $x'(\lambda) = H_\lambda^{(1)} w(\lambda) + H_\lambda w'(\lambda)$ (see footnote (14)). By (4.9) and the above results $x(\lambda)$ satisfies (4.7), and $x(0) = 0$. In case A_λ is the constant transformation A, and we take $H_\lambda = B_\lambda = e^{\lambda A}$, then $H_\lambda^{-1} = e^{-\lambda A}$ and the solution is

$$x(\lambda) = e^{\lambda A} \int_0^\lambda e^{-tA} z(t)\, dt. \tag{4.15}$$

As we shall see in §5 there is particular interest in linear transformations of finite norm in a Hilbert space $\mathcal{H}$.[16] If $\mathcal{F}$ is the class of all such transformations A of finite norm $N(A)$ $\mathcal{F}$ itself is a Hilbert space with a suitably defined inner product (A, B) such that $N(A) = (A, A)^{\frac{1}{2}}$. Moreover $\| Ax \| \leqq N(A) \| x \|$ where $\| x \| = (x, x)^{\frac{1}{2}}$ is the norm in $\mathcal{H}$. Concerning transformations of finite norm which depend on a parameter λ we have the following theorem.

THEOREM 4.3. *For each λ in Δ let A_λ be a member of $\mathcal{F}$. Let A_λ be analytic in Δ in the sense of Definition* 1.2, *and let $N(A_\lambda) < K(\Delta_0)$ in Δ_0.*[8] *Then A_λ, as a function on Δ to $\mathcal{F}$, is analytic in Δ. In particular, the transformation $A_\lambda^{(1)}$ defined by $A_\lambda^{(1)} x = \dfrac{d}{d\lambda}[A_\lambda x]$ (differentiation according to the norm in $\mathcal{H}$) is of finite norm, and*

$$\lim_{\lambda \to \lambda_0} N\left[\frac{A_\lambda - A_{\lambda_0}}{\lambda - \lambda_0} - A_{\lambda_0}^{(1)}\right] = 0.$$

PROOF: We have merely to express A_λ, A_{λ_0}, $A_{\lambda_0}^{(1)}$ by Cauchy's integral formula and make use of the formula[16]

$$N(A) = \left(\sum_{i=1}^{\infty} \| A\varphi_i \|^2\right)^{\frac{1}{2}} \tag{4.16}$$

where $\{\varphi_n\}$ is a complete orthonormal set in $\mathcal{H}$, together with the fact that $N(A_\lambda) < K(\Delta_0)$ in Δ_0. We shall omit the details.

If $\mathcal{H}$ is l_2, A_λ is defined by a matrix $\{a_{ik}(\lambda)\}$ such that each $a_{ik}(\lambda)$ is analytic in Δ, and $N^2(A_\lambda) = \sum_{i,k} | a_{ik}(\lambda) |^2$ is finite[16] for each λ, and bounded in Δ_0.[8] If $\mathcal{H}$ is L_2 the space $\mathcal{F}$ is equivalent to the space of kernels $K(s, t)$ of Hilbert-Schmidt type[16] defined almost everywhere in the square $0 \leqq s \leqq 1, 0 \leqq t \leqq 1$, with

$$\| K \|^2 = N^2(A) = \int_0^1 \int_0^1 | K(s, t) |^2 \, ds\, dt \quad \text{if} \quad Ax = \int_0^1 K(s, t) x(t)\, dt.$$

If A_λ is to have the properties described in Theorem 4.3 it is not difficult to see, in view of Theorem 3.2, that the corresponding kernel $K_\lambda(s, t)$ must be almost everywhere equal, for each λ, to a function $\bar{K}(s, t, \lambda)$ which is analytic in Δ for each pair s, t except on a fixed, two-dimensional set of measure zero, and such that

$$\int_0^1 \int_0^1 | \bar{K}(s, t, \lambda) |^2 \, ds\, dt < K(\Delta_0) \quad \text{in} \quad \Delta_0.^8$$

[16] For the definition of such transformations see Stone, loc. cit. p. 66. See also pp. 67–69 and p. 101 for matters to which we shall refer later.

5. Resolvents

Our purpose in this section is to investigate the properties of the inverse T_λ^{-1} of a linear transformation T_λ on E to E which is analytic in Δ. It is convenient to define $A_\lambda = I - T_\lambda$ (I the identity transformation) so that T_λ may be written in the form $T_\lambda = I - A_\lambda$. We recall that T_λ^{-1} exists if and only if $T_\lambda x = 0$ implies $x = 0$.

DEFINITION 5.1. *By the resolvent set Σ of A_λ with respect to the region Δ is meant the set of all points of Δ for which T_λ^{-1} exists, is linear, and defined throughout E. If λ is in the resolvent set and we write $T_\lambda^{-1} = I - R_\lambda$, the linear transformation R_λ is called the resolvent of A_λ.*

THEOREM 5.1. *The resolvent set Σ is either null, or a non-empty open subset of Δ. If it is not empty the resolvent R_λ is analytic in Σ.*

PROOF: It is possible that Σ may be the null set.[17] On the other hand, there are certainly cases in which it need not be null. Suppose that Σ is not the null set, and let λ_0 be a point of Σ. Denote by $d(\lambda_0)$ the distance from λ_0 to the nearest point of the boundary of Δ. Then if $M = M(\lambda_0)$ is the constant of (2.5) the resolvent set Σ contains all points λ for which

$$|\lambda - \lambda_0| < \text{Min}\left\{d(\lambda_0), \frac{1}{M(1 + M\|T_{\lambda_0}^{-1}\|)}\right\}. \tag{5.1}$$

To prove this we start from the identity

$$T_\lambda = T_{\lambda_0}(I - T_{\lambda_0}^{-1}(A_\lambda - A_{\lambda_0})). \tag{5.2}$$

Now if λ satisfies (5.1) we have, by (2.3) and (2.5),

$$\|(A_\lambda - A_{\lambda_0})x\| \leqq \sum_{n=1}^{\infty} |\lambda - \lambda_0|^n M^{n+1}\|x\| = \frac{M^2|\lambda - \lambda_0|}{1 - M|\lambda - \lambda_0|}\|x\| < \frac{\|x\|}{\|T_{\lambda_0}^{-1}\|}.$$

Hence $\|T_{\lambda_0}^{-1}(A_\lambda - A_{\lambda_0})x\| \leqq k(\lambda)\|x\|$, where $0 \leqq k(\lambda) < 1$. As is well known, we then have

$$\begin{gathered}(I - T_{\lambda_0}^{-1}(A_\lambda - A_{\lambda_0}))^{-1} = I + H_\lambda + H_\lambda^2 + --, \\ H_\lambda = T_{\lambda_0}^{-1}(A_\lambda - A_{\lambda_0}).\end{gathered} \tag{5.3}$$

This inverse is defined and linear throughout E. From (5.2) we can then conclude the same of T_λ^{-1}. In fact, $T_\lambda^{-1} = (I - H_\lambda)^{-1}T_{\lambda_0}^{-1}$.

It remains to prove that T_λ is analytic in Σ. It clearly suffices to prove this for $(I - H_\lambda)^{-1}$ at the point λ_0. Now H_λ is analytic in Δ, and it is not difficult to see that the same is true of H_λ^n.[18] Also

$$\|H_\lambda x\| \leqq \frac{\|T_{\lambda_0}^{-1}\|M^2|\lambda - \lambda_0|}{1 - M|\lambda - \lambda_0|}\|x\| = k(\lambda)\|x\|$$

---

[17] See the example given by Tamarkin, Annals of Mathematics, vol. 28 (1926) p. 128.

[18] The formula $\frac{d}{d\lambda}H_\lambda^2 x = H_\lambda H_\lambda^{(1)}x + H_\lambda^{(1)}H_\lambda x\left(H_\lambda^{(1)}x = \frac{d}{d\lambda}H_\lambda x\right)$ is readily established with the aid of (2.3) and (2.5). It may also be proved as corollary of a fundamental theorem about Fréchet differentials. See theorem 2 of the author's paper referred to in footnote (3).

and $k(\lambda) \leqq \rho < 1$ if

$$|\lambda - \lambda_0| < \text{Min}\left\{\rho\, d(\lambda_0), \frac{\rho}{M(1 + M \,||\, T_{\lambda_0}^{-1} \,||)}\right\}.$$

Then $|| H_\lambda^n x || \leqq \rho^n || x ||$ and so the series (5.3) converges uniformly for all λ in the neighborhood of λ_0 and $|| x || \leqq 1$, thus defining a linear transformation depending analytically on λ. Evidently the resolvent set cannot be empty unless $|| A_\lambda || \geqq 1$ in Δ.

It is known that if $A_\lambda = \lambda A$ where A is a fixed completely continuous transformation, the resolvent R_λ is 'meromorphic'—that is it has only isolated singularities in the finite plane, and if λ_0 is such a singular point R_λ may be written, in the neighborhood of λ_0, in the form $R_\lambda = (\lambda - \lambda_0)^{-n} S_\lambda$, where S_λ is linear, and analytic at λ_0, and n is a positive integer.[19] This property subsists for the resolvent (if it exists) of a transformation A_λ of finite norm in a Hilbert space $\mathfrak{H}$ if the conditions of Theorem 4.3 are satisfied. Such a transformation is completely continuous.

THEOREM 5.2. *Let $\mathfrak{H}$ be a Hilbert space; let A_λ be a linear transformation satisfying the conditions of Theorem* 4.3. *Then if the resolvent R_λ of A_λ exists it is meromorphic in Δ in the sense described above.*

PROOF: We must first state two lemmas about completely continuous transformations.[20]

LEMMA 4. *If A is a completely continuous linear transformation on E to E, and $(I - A)^{-1} \equiv I - R$ exists, then $A + R = AR = RA$. Conversely, if R is a linear transformation satisfying either of the conditions*

(i) $A + R = AR$
(ii) $A + R = RA$

then $(I - A)^{-1}$ exists and is precisely $I - R$.

This lemma is an immediate consequence of the Riesz theory. We recall that if $(I - A)^{-1}$ exists it is necessarily linear and defined throughout E. Also, the product of two linear transformations is completely continuous if at least one of the original transformations is completely continuous.

LEMMA 5. *If A is a completely continuous linear transformation on E to E a necessary and sufficient condition that $(I - A)^{-1}$ exist is that $(I - A^m)^{-1}$ exist for some integer $m > 1$.*

PROOF: The condition is necessary, for supposing that $(I - A^m)^{-1}$ does not exist for any $m = 2, 3, \cdots$ we can, for each $m > 1$ find an $x_m \neq 0$ in E such that $A^m x_m = x_m$. Let $\epsilon_1, \cdots, \epsilon_m$ be the m roots of the equation $\epsilon^m = 1$, and put

$$Y_{m,i} = x_m + \epsilon_i A x_m + \cdots + \epsilon_i^{m-1} A^{m-1} x_m$$

[19] T. H. Hildebrandt, Bulletin of the American Mathematical Society, vol. 37 (1931) pp. 197–198.

[20] The facts upon which we rely are all contained in the paper of F. Riesz, Acta Mathematica, vol. 41 (1918) pp. 71–98. A summary of what is essential will be found in the reference listed in footnote (19).

then

$$\epsilon_i A y_{m,i} = \epsilon_i A x_m + \epsilon_i^2 A^2 x_m + \cdots + \epsilon_i^m A^m x_m = y_{m,i}.$$

But $\sum_i^m y_{m,i} = m x_m \neq 0$ since $\epsilon_1^p + \cdots + \epsilon_m^p = 0$ if $1 \leqq p < m$. Hence for at least one i, say i_m, $y_{m,i_m} \neq 0$. Moreover $\epsilon_{i_m} \neq 1$, since $(I - A)^{-1}$ exists. It follows that $(I - \lambda A)^{-1}$ fails to exist for an infinite number of values of λ on the unit circle $|\lambda| = 1$. This is impossible, since A is completely continuous. Hence the necessity of the condition is established.

To prove the sufficiency suppose that $(I - A^m)^{-1} = I - R_m$ $(m > 1)$. Then R_m is completely continuous, and $A^m + R_m = R_m A^m$ by lemma 4. If we define

$$R = -(A + \cdots + A^{m-1}) + R_m + R_m(A + \cdots + A^{m-1}), \tag{5.4}$$

a direct calculation shows that $A + R = RA$. Hence, again by lemma 4, the inverse $(I - A)^{-1}$ exists and is precisely $I - R$.

Turning now to the proof of the theorem we see, by virtue of lemma 5 that if λ_0 is in the resolvent set of A_λ it is also in the resolvent set of A_λ^m for some $m > 1$. We shall find the resolvent of A_λ^m and then calculate the resolvent of A_λ by (5.4). Now A_λ^m, as the iterate of a transformation of finite norm, has the following properties:

$$\begin{aligned} &N(A_\lambda^m) \leqq [N(A_\lambda)]^m, \\ &\sum_{i=1}^{\infty} |(A_\lambda^m \varphi_i, \varphi_i)| \leqq \tfrac{1}{2}[N^2(A_\lambda) + N^2(A_\lambda^{m-1})], \end{aligned} \tag{5.5}$$

where $\{\varphi_n\}$ is a complete orthonormal set in $\mathcal{H}$, as may be verified by elementary calculations. We now seek to determine a transformation $R_{m,\lambda}$ such that $A_\lambda^m + R_{m,\lambda} = A_\lambda^m R_{m,\lambda}$. For this purpose we define, using a fixed complete orthonormal set $\{\varphi_n\}$, the quantities

$$\begin{aligned} a_{ij} &= a_{ij}(\lambda) = (A_\lambda^m \varphi_j, \varphi_i), && \lambda \in \Delta, \\ b_i &= b_i(y, \lambda) = (A_\lambda^m y, \varphi_i), && i, j = 1, 2, \cdots. \end{aligned} \tag{5.6}$$

The a_{ij}'s are analytic in Δ, and $b_i(y, \lambda)$ is a linear functional which is analytic in Δ for each y. Moreover, by (5.5) and other well known relations we have

$$\begin{aligned} &\sum_{i,j=1}^{\infty} |a_{ij}|^2 = N^2(A_\lambda^m), \\ &\sum_{i=1}^{\infty} |a_{ii}| \leqq \tfrac{1}{2}[N^2(A_\lambda) + N^2(A_\lambda^{m-1})], \\ &\sum_{i=1}^{\infty} |b_i(y, \lambda)|^2 = \|A_\lambda^m y\|^2 \leqq N^2(A_\lambda^m) \|y\|^2, \end{aligned} \tag{5.7}$$

so that all the quantities on the left are bounded in sub-regions of the type Δ_0,[8] when y is fixed. We can then apply the theory of absolutely convergent infinite determinants to obtain the solution of the system

$$\sum_{j=1}^{\infty} a_{ij} r_j - r_i = b_i, \qquad i = 1, 2, \cdots, \tag{5.8}$$

in the infinite set of unknowns $(r_1, r_2, \cdots)$.[21] But the system (5.8) is equivalent to the equation $A_\lambda^m + R_{m,\lambda} = A_\lambda^m R_{m,\lambda}$ if $R_{m,\lambda}$ is a linear transformation and

$$r_i = r_i(y, \lambda) = (R_{m,\lambda} y, \varphi_i). \tag{5.9}$$

Since we know that $(I - A_\lambda^m)^{-1}$ exists if $\lambda = \lambda_0$ it follows, by the von Koch theory,[21] that $D(\lambda_0) \neq 0$ where $D(\lambda)$ is the determinant of the system (5.8). It is then easily shown that except at the zeros of $D(\lambda)$ a unique solution $r_i(y, \lambda)$ of (5.8) exists; each $r_i(y, \lambda)$ is a linear functional, and for fixed y, is meromorphic in Δ, its only singularities being at the zeros of $D(\lambda)$. Also, as shown by Tamarkin (see reference in footnote (21), p. 135).

$$\sum_{i=1}^{\infty} |r_i(y, \lambda)|^2 \leqq \frac{e^{F(y,\lambda)}}{|D(\lambda)|},$$

$$F(y, \lambda) = N^2(A_\lambda^m) + 2 \sum_{i=1}^{\infty} |(A_\lambda^m \varphi_i, \varphi_i)| + \|A_\lambda^m y\|^2. \tag{5.10}$$

Thus we infer that the resolvent $R_{m,\lambda}$ exists (defined by (5.9)) and is meromorphic in Δ.

Applying lemma 5 and formula (5.4) the theorem is proved.

PRINCETON UNIVERSITY.

---

[21] See J. Tamarkin, loc. cit. pp. 130–136.

ANNALS OF MATHEMATICS
Vol. 39, No. 3, July, 1938

ARITHMETIC AND IDEAL THEORY OF COMMUTATIVE SEMIGROUPS

BY A. H. CLIFFORD

(Received January 24, 1938)

A set S in which a multiplication ab is defined is called a *semigroup* if this multiplication is associative and commutative, if an identity element is present in S, and if the cancellation law holds: $ab = ac$ implies $b = c$. For example, the non-zero elements of a domain of integrity constitute a semigroup under multiplication. The notions of divisibility and irreducible elements are defined in the usual way. A semigroup S will be called an *arithmetic* if every element of S is uniquely decomposable into irreducible elements. The problem is *to give necessary and sufficient conditions that a semigroup must satisfy in order that it can be embedded in an arithmetic.* Any arithmetic Σ containing a semigroup S will be called an *ideal arithmetic* of S.

This is an exact formulation of the problem of "restoring unique decomposition by the adjunction of ideal elements." The statement of the problem is in no way altered if we are considering a domain of integrity D rather than a semigroup. In that case we endeavor to embed the multiplicative semigroup S of D in an arithmetic Σ, or rather the semigroup $\bar{S}$ of principal ideals in D, which may be regarded as arising from S by identifying elements which divide each other. In Dedekind's theory of algebraic numbers, Σ is the set of all integral ideals in D, and it should be observed that Σ is not again a ring but only a semigroup—we can multiply two ideals but we cannot add them! This observation should make it clear that the problem is actually one of multiplication alone.[1]

The general solution of the problem has been effected by means of what Krull calls *v-ideals*.[2] A subset $\mathfrak{a}$ of a domain of integrity D is a v-ideal if it contains every element of D which is divisible by all common divisors of $\mathfrak{a}$ in the quotient-field of D. (Every v-ideal is also a Dedekind ideal, but not in general conversely.) Since this definition involves only the notion of divisibility, it can be applied to a semigroup S; of course we replace "quotient-field of D" by "quotient-group of S," the group of all formal quotients of elements

[1] The idea of dealing with semigroups instead of rings is by no means new. See for example R. Dedekind, *Über Zerlegungen von Zahlen durch ihre grössten gemeinschaftlichen Teiler*, Ges. Werke XXVIII, Bd. 2, p. 126.

[2] For a description of the theory of v-ideals and a full bibliography see W. Krull, *Ideal theorie*, Ergebnisse der Math., vol. 4, no. 3, 1935, §6.

of S. v-ideals have all the usual properties of Dedekind ideals. They were discovered independently by van der Waerden[3] and Arnold.[4]

Let us recall, without explicit statement, E. Noether's well-known conditions[5] on a domain of integrity D that the Dedekind ideals of D be an arithmetic:

I. Divisor-chain condition for ideals.
II. Weak multiple-chain condition for ideals.
III. Integral closure.

Van der Waerden showed that if I and III hold then the v-ideals of D are an arithmetic. Thus v-ideals will certainly succeed in restoring unique decomposition in D if the Dedekind ideals succeed, and may succeed where Dedekind ideals fail, namely for domains of integrity satisfying I and III but not II. But we still have no assurance, if v-ideals fail, that some further type of ideal cannot be found which will succeed.

This question is to a large extent settled by Arnold's work. An ideal arithmetic Σ of a semigroup S will be called *normal* if the following two conditions are satisfied:

N_1. If an element a of S divides an element b of S relative to Σ, then a divides b relative to S.

N_2. Every element of Σ is expressible as the greatest common divisor of a finite set of elements of S.

Arnold showed that *if S admits any normal ideal arithmetic Σ, then the semigroup $\mathfrak{S}$ of v-ideals in S is also a normal ideal arithmetic of S, and in fact Σ and $\mathfrak{S}$ are isomorphic.*

The next question which naturally arises is to find conditions on S which are necessary and sufficient in order that $\mathfrak{S}$ be an arithmetic. An answer to this was found by Prüfer:[6]

I′. Divisor-chain condition for v-ideals.
III′. Complete integral closure.

(For the definition of complete integral closure see Krull, l.c., p. 103, or §5 below. Prüfer gives a condition equivalent to III′. In the case of a domain of integrity, van der Waerden's conditions are sufficient, since I implies I′, and I and III together imply III′. III is necessary since III′ implies III, but I is not necessary.)

---

[3] B. L. van der Waerden, *Zur Produktzerlegung der Ideale in ganz abgeschlossenen Ringe*, Math. Ann. 101, 1929, pp. 293–308. This was reworked by Artin and presented in his form in van der Waerden's *Moderne Algebra*, vol. 2, §103, pp. 105–109. In Artin's treatment two (ordinary) ideals are "quasi-equal" if they generate the same v-ideal.

[4] I. Arnold, *Ideale in kommutativen Halbgruppen*, Rec. Math. Soc. Math. Moscow 36, 1929, pp. 401–407.

[5] E. Noether, *Abstrakter Aufbau der Idealtheorie in algebraischen Zahl- und Funktionenkörper*, Math. Ann. 96, 1927, pp. 26–61.

[6] H. Prüfer, *Untersuchungen über die Teilbarkeitseigenschaften in Körpern*, Crelle 168, 1932, p. 22.

The final question to be considered is the removal of the conditions N_1 and N_2. It is very easy to show that if a semigroup S admits an ideal arithmetic satisfying merely N_1, i.e. one which "preserves the division relation in S," then S satisfies I′ and III′, and hence $\mathfrak{S}$ is an arithmetic. For domains of integrity D, this is essentially the theorem (Krull, l.c., p. 119) that D is a "finite discrete principal order" if and only if its v-ideals constitute an arithmetic. N_1 is even more easily disposed of, but as a matter of fact it should by rights be incorporated in the notion of an ideal arithmetic; for if Σ does not preserve the division relation in S it is of no use in studying divisibility properties in S.

To summarize, if a semigroup S satisfies the conditions I′ and III′ then it admits a normal ideal arithmetic, namely $\mathfrak{S}$, and $\mathfrak{S}$ is essentially the only normal ideal arithmetic of $\mathfrak{S}$. If it does not satisfy them, then it cannot admit any ideal arithmetic preserving the division relation in S. I believe these results justify the assertion that v-ideals solve the main problem. We could hardly have expected that Dedekind ideals would be so successful, since the operation of addition plays a very essential rôle in their definition, but does not enter at all into the problem under consideration.

The purpose of the present paper is to study the arithmetic of semigroups in which the cancellation law is not postulated, hence applicable to rings containing divisors of zero. (Semigroups for which the cancellation law holds will be called *regular*.) In §1 elementary criteria are given in order that a semigroup shall be an arithmetic. Ideals are introduced in §2. In §3 the criteria of §§1 and 2 are applied to rings. Arnold's theorem is proved in §4. Finally, for the sake of completeness, regular semigroups are taken up in §5.

This paper is a revision and expansion of a thesis (California Institute of Technology, 1933) written under Professors E. T. Bell and Morgan Ward, and I wish to take this opportunity to express my appreciation of their guidance and inspiration.

1. Elementary criteria for unique decomposition

By a *semigroup* we shall mean a set S of elements $a, b, c, \cdots$ which is closed under a commutative and associative binary operation:

$$ab = ba,$$

$$a(bc) = (ab)c,$$

for all a, b, c in S. For example, every commutative ring is a semigroup under multiplication. We shall also assume that S contains an *identity element*, i.e. an element 1 with the property that

$$1 \cdot a = a \cdot 1 = a$$

for all a in S. This is not a vital assumption, for if no such element were present we could easily adjoin one.

We say that a *divides* b, and write $a \mid b$, if there exists c in S such that $ac = b$;

a is a *divisor* or *factor* of b, and b is a *multiple* of a. $a \nmid b$ shall mean that a does not divide b. If $a \mid b$ and $b \mid a$ we say that a and b are *associate*, and write $a \sim b$. A divisor of the identity element 1 is called a *unit*; the set of all units in S is a group. If $a \mid b$ but $b \nmid a$, then a is a *proper divisor* of b, and we write $a \parallel b$.

An element of S having a proper divisor which is not a unit is called *reducible*, and otherwise *irreducible*. An element of S which can be expressed as a product of a finite number of proper divisors of itself is called *decomposable*, otherwise *indecomposable*. An irreducible element is evidently indecomposable, but not always conversely. An element p of S is called *prime* if from $p \mid ab$ we can always infer $p \mid a$ or $p \mid b$, and *completely prime* if, for every positive integer n, $p^n \mid ab$ implies $p^n \mid a$ or $p \mid b$. If a prime divides a product of a finite number of elements of S, it must divide one of the factors, and so is necessarily indecomposable; it may, however, be reducible.

Let us form the successive powers $a, a^2, a^3, \cdots$ of an element a of S. There may or may not exist a power a^r of a which is divisible by a higher power a^m $(m > r)$. If there does, let r be the least such power, and suppose $a^m \mid a^r$ $(m > r)$. Then $a^{r+1} \mid a^m$, whence $a^{r+1} \mid a^r$, $a^{r+1} \sim a^r$. On multiplying successively by a we see that $a^n \sim a^r$ for every $n \geqq r$. Hence either (i) all the powers of a are non-associate, or (ii) there is a number r such that the first r powers of a are non-associate, while all higher powers are associate to a^r. In case (i), we say that a has *infinite index*; in case (ii), we say that a has *index* r. The index of an element is the number of non-associate powers of that element.

We shall say that S has the property of *unique decomposition into irreducible elements*, or briefly that S is *an arithmetic*, if

(i) every element a of S is expressible in the form

$$a = ep_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}, \tag{1.1}$$

where $p_1, p_2, \cdots, p_r$ are non-associate irreducible elements of S, $\alpha_1, \alpha_2, \cdots, \alpha_r$ are positive integers, and e is a unit; and if

(ii) from the existence of two such decompositions,

$$a = ep_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r} = e' q_1^{\beta_1} q_2^{\beta_2} \cdots q_s^{\beta_s}, \tag{1.2}$$

we can always infer that $r = s$ and, by suitable numeration, that $p_i \sim q_i$ and $p_i^{\alpha_i} \sim q_i^{\beta_i}$ $(i = 1, 2, \cdots, r)$. This last evidently means that either α_i and β_i are equal or that neither of them is less than the index of p_i (and q_i).

The associate relation $\sim$ is evidently an equivalence relation (reflexive, symmetric, and transitive), and if $a \sim a'$, $b \sim b'$, then $ab \sim a'b'$. Hence if we denote by $\bar{a}$ the class of all elements of S associate to a we can define the product of two classes $\bar{a}$ and $\bar{b}$ unambiguously by

$$\bar{a} \cdot \bar{b} = \overline{ab}.$$

The classes of associate elements of S constitute a semigroup $\bar{S}$ homomorphic with S under the mapping $a \to \bar{a}$ ($a \to \bar{a}$, $b \to \bar{b}$ imply $ab \to \bar{a}\bar{b}$). $\bar{S}$ has the property that if $\bar{a}$ and $\bar{b}$ are associate in $\bar{S}$ then they are equal. Any semigroup having this property will be called *reduced*, and $\bar{S}$ will be called the *reduced semigroup* of S. S and $\bar{S}$ have the same structure as far as those properties are concerned which depend only on division, for $a \mid b$ in S if and only if $\bar{a} \mid \bar{b}$ in $\bar{S}$. Hence $a \parallel b$ if and only if $\bar{a} \parallel \bar{b}$, p is irreducible or prime if and only if $\bar{p}$ is likewise, etc. The notion of decomposability is, however, not definable in terms of the division relation alone; if a in S is decomposable then $\bar{a}$ will be decomposable in $\bar{S}$, but not always conversely.

The definition of unique decomposition becomes a great deal simpler for a reduced semigroup S. Since the identity element 1 is the only unit in S, we may drop the factor e in (1.1). We can assume that no α_i in (1.1) exceeds the index of the corresponding irreducible p_i, for if p is an element of index r then $p^n = p^r$ for all $n \geqq r$. With this restriction we obtain absolute unicity in (ii) to within a permutation of the factors, i.e. by suitable numeration $p_i = q_i$ and $\alpha_i = \beta_i$.

If a (non-reduced) semigroup S is an arithmetic, then evidently so also is $\bar{S}$, but the converse need not be true; for if $\bar{S}$ is an arithmetic, all we know is that every element of S is *associate* to a power product of irreducibles, and not actually *equal* to such a power product multiplied by a unit. It is for this reason that we put up with a few extra complexities rather than deal only with reduced semigroups from the beginning. We shall see that this unpleasant situation cannot arise if S is a ring (Theorem 3.2).

THEOREM 1.1: *A semigroup S is an arithmetic if and only if it satisfies the following three conditions*:

A. (*Divisor-chain condition.*) *There exists in S no infinite sequence $a_1, a_2, \cdots, a_n, \cdots$ such that $a_{n+1} \parallel a_n$ for all positive integers n.*

B. *Every reducible element of S is decomposable.*

C. *Every irreducible element of S is completely prime.*

PROOF: If every proper divisor of an element a of S is expressible as a finite product of indecomposable elements, then the same is true of a. Hence if there exists an element a_1 of S which is not so expressible, then a_1 has a proper divisor a_2 with the same property, a_2 has a proper divisor a_3 with the same property, etc. Therefore if S satisfies condition **A** we conclude that every element a of S is expressible as a finite product,

$$a = p_1 p_2 \cdots p_r,$$

of indecomposable elements of S. If S satisfies **B** also, then $p_1, p_2, \cdots, p_r$ are irreducible.

To get this in the desired form (1.1) we note that if p and q are associate irreducible elements of S, then either they differ only by a unit factor or else they are both of index one. For since $q \mid p$ we may write $p = cq$. Since p is irreducible, c is either a unit or associate to p; in the latter case, from $c \sim p$

and $q \sim p$ we infer $p \sim p^2$. Hence a product $p_1 p_2 \cdots p_k$ of k associate irreducibles p_i of index greater than one can be put in the form ep_1^k, where e is a unit. If, on the other hand, the p_i all have index one, then the whole product is itself an irreducible element.

Condition **C** assures the unicity. For if (1.2) holds, then p_1 divides the product on the right and so by **C** must divide one of the factors (other than e'), say q_1. Since q_1 is irreducible, this means that $p_1 \sim q_1$. Similarly p_2 must be associate to one of the q_i, and this q_i cannot be q_1 since this would imply $p_1 \sim p_2$; we may therefore call it q_2. Continuing thus we see that $r \leqq s$, and hence by symmetry $r = s$, and that $p_i \sim q_i$ $(i = 1, 2, \cdots, r)$ in the new enumeration of the q's. Since moreover

$$p_1^{\alpha_1} \mid q_1^{\beta_1}(q_2^{\beta_2} \cdots q_r^{\beta_r}), \qquad p_1 \nmid q_2^{\beta_2} \cdots q_r^{\beta_r},$$

it follows from **C** that $p_1^{\alpha_1} \mid q_1^{\beta_1}$, and hence by symmetry $p_1^{\alpha_1} \sim q_1^{\beta_1}$.

Proceeding now to the converse, we note first that **B** is immediate. For if a reducible element a of S has the decomposition (1.1) then evidently $e, p_1, \cdots, p_r$ are all proper divisors of a. Now the reduced semigroup $\bar{S}$ of S is also an arithmetic, and if $\bar{S}$ has the properties **A** and **C** then S has them also. Hence we need only show that a reduced arithmetic S has the properties **A** and **C**.

Let

$$a = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$$

be any element of S, where the p_i are distinct irreducibles. It is readily seen that all divisors of a are described by the elements

$$p_1^{\xi_1} \cdots p_r^{\xi_r} \qquad (0 \leqq \xi_i \leqq \alpha_i).$$

Since these are finite in number, it follows that no element of S can have an infinite number of divisors, from which **A** follows immediately.

To show **C**, suppose that p is an irreducible element of S, $p^n \mid ab$ (n any positive integer), but $p \nmid b$. Then c exists in S such that $p^n c = ab$. We may assume furthermore that n does not exceed the index of p. On thinking of a, b, and c decomposed into irreducibles, it follows from the unicity of the decomposition of ab, and the fact that the irreducible p does not occur in the decomposition of b, that p must occur in that of a with multiplicity not less than n.

An element d of S is called a *greatest common divisor* (GCD) of a subset A of S if (i) d is a common divisor of A, i.e. divides every element of A, and (ii) d is divisible by every other common divisor of A. If such a d exists it is unique to within associate elements. If A is finite, and consists of the elements $a_1, a_2, \cdots, a_m$ of S, we shall write

$$d \sim (a_1, a_2, \cdots, a_m).$$

If every pair of elements a, b of S has a GCD then by a simple induction one sees readily that every finite set of elements of S has a GCD, and one may insert parentheses at will in $(a_1, a_2, \cdots, a_m)$.

THEOREM 1.2: *Condition C of Theorem* 1.1 *may be replaced by the following two conditions*:

D. *Every pair of elements* a, b *of* S *has a GCD* (a, b).

E. *For every three elements* a, b, c *of* S,

$$a(b, c) \sim (ab, ac).$$

PROOF: We first note three simple consequences of **D** and **E**.

$$(1.3) \qquad (a_1, \cdots, a_m)(b_1, \cdots, b_n) \sim (a_1b_1, \cdots, a_ib_j, \cdots, a_mb_n),$$

where all the mn terms a_ib_j $(i = 1, \cdots, m; j = 1, \cdots, n)$ appear on the right.

(1.4) If $(a, b) \sim 1$ and $(a, c) \sim 1$, then $(a, bc) \sim 1$.

(1.5) If $a \mid bc$ and $(a, b) \sim 1$, then $a \mid c$.

To establish **C**, let p be an irreducible element of S, and suppose that $p^n \mid ab$, $p \nmid b$. From $p \nmid b$ follows $(p, b) \sim 1$. By repeated applications of (1.4), $(p^n, b) \sim 1$, and by (1.5), $p^n \mid a$.

To show the converse, we may again assume that the arithmetic S is reduced, for **D** and **E** will hold in S if they hold in $\bar{S}$. If a, b, c are any three elements of S, we can write

$$a = p_1^{\alpha_1} \cdots p_r^{\alpha_r},$$

$$b = p_1^{\beta_1} \cdots p_r^{\beta_r},$$

$$c = p_1^{\gamma_1} \cdots p_r^{\gamma_r},$$

where some of the integers α_i, β_i, γ_i may be zero. All common divisors of a and b are described by the elements

$$p_1^{\xi_1} \cdots p_r^{\xi_r} \qquad (0 \leqq \xi_i \leqq \min(\alpha_i, \beta_i)).$$

Hence

$$d = p_1^{\delta_1} \cdots p_r^{\delta_r}, \qquad \delta_i = \min(\alpha_i, \beta_i),$$

is a GCD of a and b, which proves **D**. Since

$$\min(\alpha_i + \beta_i, \alpha_i + \gamma_i) = \alpha_i + \min(\beta_i, \gamma_i),$$

it follows that $(ab, ac) = a(b, c)$, proving **E**.

2. Ideals

A subset $\mathfrak{a}$ of a semigroup S is called an *ideal* if it contains every element b of S for which the following proposition is true: whenever two elements s and t of S are such that $s \mid at$ for all a in $\mathfrak{a}$, then we can infer that $s \mid bt$.

If A is any subset of S, the set (A) of all elements a' of S such that $s \mid at$ (all a in A) implies $s \mid a't$ is an ideal which is contained in every ideal containing A. (A) is called the ideal *generated by* A, and A is a *basis* of (A). An ideal having a finite basis of elements $a_1, a_2, \cdots, a_m$ of S may also be denoted

by $(a_1, a_2, \cdots, a_m)$; such an ideal will be called *finite*. A *principal ideal* is an ideal (a) having a single basis element a; it consists of all multiples of a.

The *ideal sum* of a set of ideals in S is defined to be the ideal generated by their class sum, i.e. the smallest ideal containing all of them. We denote the sum of a finite set of ideals $\mathfrak{a}_1, \mathfrak{a}_2, \cdots, \mathfrak{a}_m$ by $(\mathfrak{a}_1, \mathfrak{a}_2, \cdots, \mathfrak{a}_m)$. Evidently

$$((\mathfrak{a}, \mathfrak{b}), \mathfrak{c}) = (\mathfrak{a}, (\mathfrak{b}, \mathfrak{c})) = (\mathfrak{a}, \mathfrak{b}, \mathfrak{c}),$$

so that we can remove or insert parentheses at will. Note also that

$$((a_1), (a_2), \cdots, (a_m)) = (a_1, a_2, \cdots, a_m).$$

The *product* $\mathfrak{a}\mathfrak{b}$ of two ideals $\mathfrak{a}$ and $\mathfrak{b}$ is defined to be the ideal generated by the set of all products ab with a in $\mathfrak{a}$ and b in $\mathfrak{b}$. We may write $a\mathfrak{b}$ for $(a)\mathfrak{b}$. If A and B are arbitrary subsets of S, then

$$(AB) = (A)(B), \tag{2.1}$$

where AB is the set of all products ab with a in A and b in B. For (AB) is evidently contained in $(A)(B)$, and contains all products $a'b'$ with a' in (A) and b' in (B). To prove the last assertion, let s and t be any two elements of S such that $s \mid abt$ for all a in A and all b in B. Since a' is in (A), this implies $s \mid a'bt$ for all b in B. Since b' is in (B), this in turn implies $s \mid a'b't$.

In particular, for finite ideals,

$$(a_1, \cdots, a_m)(b_1, \cdots, b_n) = (a_1b_1, \cdots, a_ib_j, \cdots, a_mb_n). \tag{2.2}$$

From (2.1), the associative law

$$\mathfrak{a}\mathfrak{b}\cdot\mathfrak{c} = \mathfrak{a}\cdot\mathfrak{b}\mathfrak{c}$$

and the distributive law

$$\mathfrak{a}(\mathfrak{b}, \mathfrak{c}) = (\mathfrak{a}\mathfrak{b}, \mathfrak{a}\mathfrak{c}) \tag{2.3}$$

follow easily.

S itself is the principal ideal (1). Evidently

$$(1)\mathfrak{a} = \mathfrak{a}(1) = \mathfrak{a}$$

for every ideal $\mathfrak{a}$. Since plainly

$$\mathfrak{a}\mathfrak{b} = \mathfrak{b}\mathfrak{a},$$

it follows that the set $\mathfrak{S}$ of all ideals in S is a semigroup. Since $(a)(b) = (ab)$, and $(a) = (b)$ if and only if $a \sim b$, the set $\mathfrak{S}_0$ of all principal ideals in S is a subsemigroup of $\mathfrak{S}$ isomorphic with the reduced semigroup of S. If $\mathfrak{a} \mid \mathfrak{b}$ in $\mathfrak{S}$, i.e. if an ideal $\mathfrak{c}$ exists such that $\mathfrak{a}\mathfrak{c} = \mathfrak{b}$, then plainly $\mathfrak{a} \supseteq \mathfrak{b}$. Hence if $\mathfrak{a} \sim \mathfrak{b}$ in $\mathfrak{S}$ then $\mathfrak{a} = \mathfrak{b}$, so that $\mathfrak{S}$ is a reduced semigroup.

THEOREM 2.1: *If a semigroup S satisfies the condition* **F** *below, then every ideal in S is finite.*

F. *There exists no infinite sequence of finite ideals* $\mathfrak{a}_1, \mathfrak{a}_2, \mathfrak{a}_3, \cdots$ *in* S *such that*[7] $\mathfrak{a}_n \subset \mathfrak{a}_{n+1}$ $(n = 1, 2, 3, \cdots)$.

PROOF: Let $\mathfrak{a}$ be any ideal, and a_1 any element of $\mathfrak{a}$. If $\mathfrak{a} \neq (a_1)$, $\mathfrak{a}$ contains a_2 not in (a_1). If $\mathfrak{a} \neq (a_1, a_2)$, $\mathfrak{a}$ contains a_3 not in (a_1, a_2). Continuing thus, we construct a sequence of finite ideals $\mathfrak{a}_n = (a_1, a_2, \cdots, a_n)$ such that $\mathfrak{a}_n \subset \mathfrak{a}_{n+1}$. Hence by **F** we must have $\mathfrak{a}_n = \mathfrak{a}$ for some finite n.

THEOREM 2.2: *Condition* **C** *of Theorem* 1.1 *may be replaced by the following*: **G**. *Every ideal in* S *is a principal ideal.*

PROOF: We prove the sufficiency of **A**, **B**, and **G** by showing that **G** implies the conditions **D** and **E** of Theorem 1.2. If $(a, b) = (d)$, then it is easily verified that d is a GCD of a and b, and by (2.2)

$$(ac, bc) = (a, b)(c) = (d)(c) = (dc).$$

We prove the necessity by showing that **G** is a consequence of the conditions **A**, **D**, and **E**. From **D** it follows that every finite set of elements $a_1, \cdots, a_m$ of S has a GCD d. Since $d \mid a_i$ $(i = 1, \cdots, m)$, $(a_1, \cdots, a_m) \subseteq (d)$. To show that $(a_1, \cdots, a_m) \supseteq (d)$, suppose that s and t are any two elements of S such that $s \mid a_i t$ $(i = 1, \cdots, m)$. By **E**, dt is a GCD of $a_1 t, \cdots, a_m t$, and hence $s \mid dt$. This shows that every finite ideal is principal. But then **A** is equivalent to the condition **F** of Theorem 2.1. Every ideal in S is therefore finite, and hence principal.

If $\mathfrak{a}$ and $\mathfrak{b}$ are ideals, the set of all elements c in S such that $ac \subseteq \mathfrak{b}$ is an ideal $\mathfrak{c}$ which we denote by $\mathfrak{b}:\mathfrak{a}$ and call the *quotient ideal* of $\mathfrak{b}$ by $\mathfrak{a}$. Condition **K** in the following theorem is a strengthening of the obvious relations

$$\mathfrak{a}:\mathfrak{b} \supseteq \mathfrak{a} \supseteq (\mathfrak{a}:\mathfrak{b})\cdot\mathfrak{b}.$$

Condition **H** requires that the relations "containing" and "dividing" shall be synonymous.

THEOREM 2.3: *The two conditions* **H** *and* **J** *below* (*on a semigroup* S) *are equivalent to the two conditions* **F** *and* **K**:

H. $\mathfrak{a} \supseteq \mathfrak{b}$ *implies* $\mathfrak{a} \mid \mathfrak{b}$($\mathfrak{a}$, $\mathfrak{b}$ *ideals in* S).

J. *The semigroup* $\mathfrak{S}$ *of ideals in* S *is an arithmetic.*

F. *Chain condition for finite ideals* (*Theorem* 2.1).

K. *If* $\mathfrak{a} \subset \mathfrak{b} \subset (1)$, *then* $\mathfrak{a}:\mathfrak{b} \supset \mathfrak{a}$ *and* $\mathfrak{a} = (\mathfrak{a}:\mathfrak{b})\cdot\mathfrak{b}$.

PROOF: Assume **H** and **J**. Applying condition **A** of Theorem 1.1 to $\mathfrak{S}$, we see that there can exist no infinite sequence of ideals $\mathfrak{a}_1, \mathfrak{a}_2, \cdots$ such that $\mathfrak{a}_{n+1} \,||\, \mathfrak{a}_n$. By **H**, $\mathfrak{a}_{n+1} \,||\, \mathfrak{a}_n$ is equivalent to $\mathfrak{a}_{n+1} \supset \mathfrak{a}_n$ which proves **F**.

If $\mathfrak{a} \subset \mathfrak{b} \subset (1)$ then $\mathfrak{b} \,||\, \mathfrak{a}$ by **H**. Since $\mathfrak{S}$ is a reduced arithmetic and $\mathfrak{b} \neq (1)$, it follows that $\mathfrak{a} = \mathfrak{b}\mathfrak{c}$, where $\mathfrak{c} \,||\, \mathfrak{a}$, i.e. $\mathfrak{a} \subset \mathfrak{c}$. But from $\mathfrak{a} = \mathfrak{b}\mathfrak{c}$ follows $\mathfrak{c} \subseteq \mathfrak{a}:\mathfrak{b}$, whence $\mathfrak{a} \subset \mathfrak{c} \subseteq \mathfrak{a}:\mathfrak{b}$ and $\mathfrak{a} = \mathfrak{b}\mathfrak{c} \subseteq \mathfrak{b}(\mathfrak{a}:\mathfrak{b})$, which proves **K**.

Conversely, assume **F** and **K**. **H** is an immediate consequence of the second part of **K**. We prove **J** by proving **A**, **B**, **D**, and **E** of Theorem 1.2 applied to $\mathfrak{S}$.

---

[7] The symbol $\subset$ means "is a proper subset of."

A follows from **H** and **F**, **B** from **H** and **K**. By **H**, the ideal sum $(\mathfrak{a}, \mathfrak{b})$ of two ideals $\mathfrak{a}$ and $\mathfrak{b}$ is the GCD of $\mathfrak{a}$ and $\mathfrak{b}$, proving **D**. **E** then follows from (2.3).

3. Applications to rings

We interrupt the main development at this point to apply the criteria of Theorems 1.1, 1.2, and 2.2 to (commutative) rings which may contain divisors of zero. An element of a ring is *regular* if it is not a divisor of zero; a regular element must be of infinite index. The main purpose of this section is to show that the condition **B** can be replaced by the following:

L. *If* $a_1, a_2, a_3, \cdots$ *is an infinite sequence of elements such that* $a_n \mid\mid a_{n+1}$ $(n = 1, 2, 3, \cdots)$, *then no element* $a \neq 0$ *is divisible by all the elements* a_n *of the sequence.*

THEOREM 3.1. *If* R *is a ring satisfying the condition* **L**, *then every reducible element* $\neq 0$ *of* R *is decomposable. Hence for rings the condition* **B** *of Theorems* 1.1, 1.2, *and* 2.2 *may be replaced by* **L**.

PROOF:[8] Let $a \neq 0$ be a reducible element of R, and b a proper divisor of a, not a unit. Then $a = bc$ with c in R. If $c \mid\mid a$ the result is established. If $c \sim a$ then $bc \sim ba$, so that $a \sim ba$. On multiplying again and again by b, we see that $a \sim b^n a$ for all positive integers n, and hence that a is divisible by each element of the infinite sequence $b, b^2, b^3, \cdots$. From **L** it follows that b must be of finite index, and hence a divisor of zero, i.e. there exists $d \neq 0$ in R such that $bd = 0$. We proceed to show that the equation

$$a = b(c + d)$$

is a decomposition of a, i.e. $c + d \mid\mid a$. Now $a \mid c$, say $af = c$, and hence if a were a divisor of $c + d$ it would divide d, say $ag = d$. But from this we would infer

$$d = ag = bcg = bafg = bfd = f \cdot 0 = 0,$$

in contradiction to $d \neq 0$.

The second part of the theorem is an immediate consequence of the first part, together with the evident necessity of **L**.

If a ring R is an arithmetic and actually contains divisors of zero, then 0 itself has a unique decomposition,

$$0 = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_n^{\alpha_n},$$

and R is essentially finite; the reduced semigroup of R has exactly $(\alpha_1 + 1)(\alpha_2 + 1) \cdots (\alpha_n + 1)$ elements. When we say that a domain of integrity is an arithmetic we naturally exclude the (indecomposable) element 0.

As an application of Theorem 3.1, we prove two simple corollaries.

---

[8] Due to A. Fränkel, *Über die Teiler der Null und die Zerlegung von Ringen*, Crelle 145, 1914, p. 150.

COROLLARY 1: *If every Dedekind ideal of a ring R is a principal ideal, and R satisfies the condition* **L** *of Theorem* 3.1, *then R is an arithmetic.*

The first of these conditions, as is well known, implies the condition **A** of Theorem 1.1. Since every ideal of the type discussed in §2 is also a Dedekind ideal, condition **G** of Theorem 2.2 holds.

COROLLARY 2: *If $\mathfrak{a}$ is any Dedekind ideal $\neq (0)$ in a domain of integrity D whose Dedekind ideals constitute an arithmetic, then the quotient-ring $D/\mathfrak{a}$ is an arithmetic.*

In the first place, every Dedekind ideal in $D/\mathfrak{a}$ is a principal ideal.[9] In the second place, there can exist no infinite sequence of ideals $\mathfrak{a}_1 \supset \mathfrak{a}_2 \supset \mathfrak{a}_3 \supset \cdots$ in D all containing $\mathfrak{a}$; this is the condition II of E Noether mentioned in the introduction. Applying this to principal ideals in $D/\mathfrak{a}$, we see that **L** is true, and the result follows from Corollary 1.

The rest of this section will be devoted to showing that if a ring R having an identity element 1 is an arithmetic, then associate elements of R differ only by a unit factor, i.e. if $a \sim b$ then $a = eb$ where e is a unit.

LEMMA 3.1: *If, in a semigroup S, every irreducible p of index one differs from its square p^2 by a unit factor, then any two associate irreducibles of S differ by a unit factor.*

PROOF: In the proof of Theorem 1.1 it was shown that any two associate irreducibles of index >1 differ by a unit factor, so we need only prove the case of index one.

We note first that if p is an idempotent irreducible ($p^2 = p$) and $px = p$, then either x is a unit or $x = p$. For if x is not a unit, then $p \mid x$, say $py = x$; then $p = px = p^2y = py = x$.

From this it follows that two associate idempotent irreducibles p and q differ by a unit factor. For if $pa = q$ and $qb = p$, then

$$pab = qb = p, \qquad qab = pa = q.$$

If ab is not a unit then $p = ab$ and $q = ab$, whence $p = q$.

Suppose now that p and q are any two associate irreducibles of index one. By hypothesis there exist units u and v in S such that $p^2u = p$ and $q^2v = q$. But then pu and qv are associate idempotent irreducibles; hence there is a unit w in S such that $pu = qvw$, or $p = q \cdot u^{-1}vw$.

THEOREM 3.2: *In a ring R having an identity element, any two associate irreducibles differ by a unit factor. If R is an arithmetic, then any two associate elements of R differ by a unit factor.*

PROOF: The second part is of course an immediate consequence of the first. By Lemma 3.1 we have only to show that an irreducible p in R of index one differs from p^2 by a unit factor.

---

[9] See for example van der Waerden, l.c., vol. 2, p. 102.

Since p is of index one, $p^2 \mid p$, say $p^2a = p$. If a is not itself a unit then $p \mid a$, say $pb = a$. Then $a = pb = p^2ab = pa \cdot pb = pa^2$. We therefore have

$$p(1 - pa) = a(1 - pa) = 0.$$

Since pa is idempotent, the same is true of $1 - pa$. Hence

$$\begin{aligned}(1 - pa + p)(1 - pa + a) &= (1 - pa)^2 + (p + a)(1 - pa) + pa \\ &= 1 - pa + pa = 1.\end{aligned}$$

Thus $u = 1 - pa + p$ is a unit, and

$$pu = p(1 - pa) + p^2 = p^2.$$

4. Normal ideal arithmetics

A reduced semigroup Σ will be called an *ideal arithmetic* of a semigroup S if (i) Σ is a reduced arithmetic and (ii) Σ contains a subsemigroup Σ_0 isomorphic with the reduced semigroup $\bar{S}$ of S. We may identify Σ_0 with $\bar{S}$ in the usual way, and simply say that $\Sigma \supseteq \bar{S}$. Since the existence or non-existence of such a Σ concerns $\bar{S}$ rather than S, we shall deal only with reduced semigroups throughout this section. The notations $a \mid_S b$ and $a \mid_\Sigma b$ will be used to indicate division relative to S and Σ respectively; unless otherwise indicated, division will always be understood to be relative to Σ. Since Σ is an arithmetic, every pair of elements α, β of Σ has by Theorem 1.2 a GCD in Σ which we now denote by $\{\alpha, \beta\}$.

An ideal arithmetic Σ of a reduced semigroup S is called *normal* if the following two conditions are satisfied:

$\mathbf{N}_1$. *Σ preserves the division relation in S, i.e. $a \mid_\Sigma b$ if and only if $a \mid_S b$ (a, b in S).*

$\mathbf{N}_2$. *Each element α of Σ is the GCD of a finite set of elements $a_1, \cdots, a_m$ of S: $\alpha = \{a_1, \cdots, a_m\}$.*

LEMMA 4.1: *If α is any element of a normal ideal arithmetic Σ of a reduced semigroup S, then there exists a finite set of elements $\xi_1, \xi_2, \cdots, \xi_r$ of Σ such that $\alpha\xi_1, \alpha\xi_2, \cdots, \alpha\xi_r$ are in S and $\{\xi_1, \xi_2, \cdots, \xi_r\} = 1$.*

PROOF: Let ρ be the GCD of the set Ξ of all elements ξ of Σ such that $\alpha\xi$ is in S. Then $\alpha\rho$ is the GCD of the set $\alpha\Xi$. But the latter is precisely the set of all multiples of α in S, so that α is the GCD of $\alpha\Xi$ by N_2. Hence $\alpha\rho = \alpha$; If $\rho = 1$ the result is evident since the GCD of any set of elements of an arithmetic is the GCD of a finite subset thereof. We proceed to show that the assumption $\rho \neq 1$ leads to a violation of the divisor-chain condition in Σ (**A** of Tbeorem 1.1).

In the first place, if $\rho \neq 1$ then it must be a proper divisor of every ξ in Ξ. for if ρ were itself in Ξ then $\alpha(= \alpha\rho)$ would be in S, and Ξ would contain 1. Hence if ξ_1 is any element of Ξ there exists a proper divisor ξ_2 of ξ_1 such that $\xi_1 = \rho\xi_2$. But then $\alpha\xi_1 = \alpha\xi_2$, so that ξ_2 is also in Ξ. Thus every element of Ξ has a proper divisor in Ξ, contrary to the divisor-chain condition.

LEMMA 4.2: *Let Σ be a normal ideal arithmetic of a reduced semigroup S, and let $a_1, a_2, \cdots, a_m$ be a finite number of elements of S. Then every element of the ideal $(a_1, a_2, \cdots, a_m)$ is divisible by $\{a_1, a_2, \cdots, a_m\}$.*

PROOF: Let $\alpha = \{a_1, \cdots, a_m\}$, and let a be any element of the ideal $(a_1, \cdots, a_m)$. We show first that if ξ is any element of Σ such that $\alpha\xi = c$ is in S, then $\alpha \mid a\xi$.

By $\mathbf{N}_2$, $\xi = \{s_1, \cdots, s_n\}$ with $s_1, \cdots, s_n$ in S, and by $\mathbf{N}_1$

$$c \mid_S a_i s_j \qquad (i = 1, \cdots, m; j = 1, \cdots, n).$$

Since a is in $(a_1, \cdots, a_m)$,

$$c \mid_S a s_j \qquad (j = 1, \cdots, n).$$

Hence c divides the GCD $\{as_1, \cdots, as_n\} = a\xi$ of the as_j. Since $\alpha \mid c$ we conclude that $\alpha \mid a\xi$.

By Lemma 4.1, there exist $\xi_1, \cdots, \xi_r$ in Σ such that $\alpha\xi_1, \cdots, \alpha\xi_r$ are in S and $\{\xi_1, \cdots, \xi_r\} = 1$. From the foregoing, α divides the elements $a\xi_1, \cdots, a\xi_r$ and hence also their GCD $\{a\xi_1, \cdots, a\xi_r\} = a$.

THEOREM 4.1: *The semigroup $\mathfrak{S}$ of ideals in a reduced semigroup S is a normal ideal arithmetic of S if and only if* (i) *$\mathfrak{S}$ is an arithmetic and* (ii) *$\mathfrak{a} \supseteq \mathfrak{b}$ implies $\mathfrak{a} \mid_{\mathfrak{S}} \mathfrak{b}$ ($\mathfrak{a}$, $\mathfrak{b}$ ideals in S).*

These are the conditions **J** *and* **H** *of Theorem* 2.3; *hence we can also say that $\mathfrak{S}$ is a normal ideal arithmetic of S if and only if* **F** *and* **K** *are satisfied.*

PROOF. Assume **J** and **H**. Since the set $\mathfrak{S}_0$ of principal ideals in S is a subsemigroup of $\mathfrak{S}$ isomorphic with S, it follows from **J** that $\mathfrak{S}$ is an ideal arithmetic of S. If $a \mid_{\mathfrak{S}} b$ (a, b in S) then $(a) \supseteq (b)$, $a \mid_S b$, and $\mathbf{N}_1$ holds. Since **J** and **H** imply **F**, every ideal $\mathfrak{a}$ has by Theorem 2.1 a finite basis, $\mathfrak{a} = (a_1, \cdots, a_m)$. From **H** it follows that $\mathfrak{a}$ is the GCD in $\mathfrak{S}$ of $a_1, \cdots, a_m$, and $\mathbf{N}_2$ holds.

Suppose conversely that $\mathfrak{S}$ is a normal ideal arithmetic of S; we need of course only prove **H**. By $\mathbf{N}_2$, every ideal $\mathfrak{a}$ is the GCD of a finite number of principal ideals $(a_1), (a_2), \cdots, (a_m)$. Since $\mathfrak{a}$ contains each of the latter, it contains the ideal $(a_1, a_2, \cdots, a_m)$. Let s and t be any two elements of S such that $s \mid_S a_i t$ $(i = 1, \cdots, m)$. Since $\mathfrak{a}(t)$ is the GCD of the ideals $(a_i)(t)$, we have $(s) \mid_{\mathfrak{S}} \mathfrak{a}(t)$ and hence $(s) \supseteq \mathfrak{a}(t)$. If a is any element of $\mathfrak{a}$, $(s) \supseteq \mathfrak{a}(t) \supseteq (a)(t)$ and hence $s \mid_S at$. Thus every element of $\mathfrak{a}$ is in $(a_1, \cdots, a_m)$, whence $\mathfrak{a} = (a_1, \cdots, a_m)$.

From Lemma 4.2 it follows that $\mathfrak{a}$ divides every element of $\mathfrak{a}$. Hence if $\mathfrak{b} \supseteq \mathfrak{a}$ then $\mathfrak{b}$ divides each of the elements a_i and therefore divides their GCD $\mathfrak{a}$.

THEOREM 4.2. *If a reduced semigroup S admits a normal ideal arithmetic Σ then the semigroup $\mathfrak{S}$ of ideals in S is also a normal ideal arithmetic of S, and in fact $\mathfrak{S}$ and Σ are isomorphic.*

The isomorphism between $\mathfrak{S}$ and Σ is established in such a way that if α in Σ corresponds to $\mathfrak{a}$ in $\mathfrak{S}$ then α is the GCD in Σ of the elements of $\mathfrak{a}$ and $\mathfrak{a}$ is the set of all multiples of α in S. In particular, the elements of S are invariant if we identify S with the semigroup $\mathfrak{S}_0$ of principal ideals.

PROOF. According to $\mathbf{N}_2$, every element α of Σ can be represented as the GCD $\{a_1, a_2, \cdots, a_m\}$ of a finite set of elements $a_1, a_2, \cdots, a_m$ of S. The

bulk of the proof will be to show that we can set up a one-to-one correspondence between Σ and the semigroup $\mathfrak{S}'$ of all finite ideals in S such that $\{a_1, a_2, \cdots, a_m\}$ and $(a_1, a_2, \cdots, a_m)$ correspond. To do this we must show that

$$\{a_1, a_2, \cdots, a_m\} = \{b_1, b_2, \cdots, b_n\}$$

if and only if

$$(a_1, a_2, \cdots, a_m) = (b_1, b_2, \cdots, b_n).$$

This will follow if we can show that $(a_1, \cdots, a_m)$ coincides with the set of all multiples of $\{a_1, \cdots, a_m\}$ in S, and hence that $\{a_1, \cdots, a_m\}$ is the GCD in Σ of all the elements of $(a_1, \cdots, a_m)$. By Lemma 4.2 every element of $(a_1, \cdots, a_m)$ is a multiple of $\{a_1, \cdots, a_m\}$; hence we have only to show the converse.

Let $\alpha = \{a_1, \cdots, a_m\}$, and let a be any multiple of α in S. Let s and t be any two elements of S such that $s \mid_S a_i t$ $(i = 1, \cdots, m)$. Since αt is the GCD of the elements $a_1 t, \cdots, a_m t$ it follows that $s \mid_\Sigma \alpha t$. Since $\alpha \mid_\Sigma a$, we conclude that $s \mid_\Sigma at$, and hence, by $\mathbf{N}_1$, $s \mid_S at$, so that a is in $(a_1, \cdots, a_m)$.

That the correspondence thus established between $\mathfrak{S}'$ and Σ is an isomorphism is evident from the multiplication laws (1.3) and (2.2). Now if $(a_1, \cdots, a_m) \supset (b_1, \cdots, b_n)$ then $\alpha \,||\, \beta$, where $\alpha = \{a_1, \cdots, a_m\}$ and $\beta = \{b_1, \cdots, b_n\}$. For $\alpha \mid \beta$ from what we have shown above, and were $\alpha = \beta$ we would have equality between the ideals. Consequently we can conclude the chain condition **F** of Theorem 2.1 for finite ideals in S from the chain condition **A** of Theorem 1.1 for elements of the arithmetic Σ. Thus every ideal in S is finite, $\mathfrak{S}' = \mathfrak{S}$, and hence $\mathfrak{S}$ and Σ are isomorphic. From this isomorphism it naturally follows that $\mathfrak{S}$ is a normal ideal arithmetic of S.

This theorem has two important aspects. In the first place it tells us that if a reduced semigroup S admits a normal ideal arithmetic of any kind whatever, then the semigroup $\mathfrak{S}$ of ideals in S—which we can always construct, at least theoretically—will serve the purpose. Thus, by Theorem, 4.1, the conditions **F** and **K** of Theorem 2.3 are necessary and sufficient conditions that S admit a normal ideal arithmetic of any kind. In the second place it tells us that if these conditions are satisfied then S has—to within isomorphism—only one normal ideal arithmetic, namely $\mathfrak{S}$; $\mathfrak{S}$ is then characterized by the conditions $\mathbf{N}_1$ and $\mathbf{N}_2$. If, in particular, the Dedekind ideals in a ring constitute a normal ideal arithmetic thereof, then they coincide with the v-ideals. Note that if S is itself an arithmetic then it is a normal ideal arithmetic of itself and hence $\mathfrak{S}$ coincides with S, i.e. every ideal is a principal ideal (Theorem 2.2).

5. Regular semigroups

A semigroup S will be called *regular* if the cancellation law holds in S, i.e. from $ab = ac$ we can infer $b = c$. The non-zero elements of a domain of integrity constitute a regular semigroup under multiplication.

Every non-unit element of a regular semigroup S has infinite index; hence in

the definition (§1) of unique decomposition we must have $\alpha_i = \beta_i$. Two associate elements of S differ only by a unit factor; hence S is an arithmetic if its reduced semigroup $\bar{S}$ is, and we could restrict ourselves to reduced semigroups from the beginning. If $b \mid\mid a$, $a = bc$, b not a unit, then $c \mid\mid a$; hence every reducible element of S is decomposable. Thus—for domains of integrity—the developments of §3 are quite unnecessary. It is readily seen that every prime element of S is completely prime, and that the condition **E** of Theorem 1.2 is a consequence of **D**. Theorems 1.1, 1.2, and 2.2 now become a good deal simpler.

THEOREM 5.1. *A regular semigroup S is an arithmetic if and only if it satisfies the divisor-chain condition* **A** *of Theorem* 1.1 *together with one of the following three conditions*:

C′. *Every irreducible element of S is prime.*

D. *Every pair of elements of S has a GCD in S.*

G. *Every ideal in S is a principal ideal.*

We can embed every regular semigroup S in a group G by constructing formal quotients ab^{-1} of elements a, b of S in the well-known way, i.e. $ab^{-1} = a'b'^{-1}$ if $ab' = a'b$, and $(ab^{-1})(cd^{-1}) = (ac)(bd)^{-1}$. If S is the multiplicative semigroup of the non-zero elements of a domain of integrity D, then G is the multiplicative group of the non-zero elements of the quotient-field of D. We call G the *quotient-group* of S.

The notion of division relative to S can be extended in the usual way to elements α, β of G, i.e. $\alpha \mid_S \beta$ if $\beta\alpha^{-1}$ is in S. A subset A of G is called *integral* if $A \subseteq S$, and *fractional* if there exists a common denominator c in S for all the elements of A, i.e. $cA \subseteq S$. The set of all elements ξ of G for which $\xi A \subseteq S$ is denoted by A^{-1}.

We may now say that an element a' of S belongs to an ideal $\mathfrak{a}$ if from $st^{-1} \mid a$ (all a in $\mathfrak{a}$) we can infer $st^{-1} \mid a'$, i.e. if a' is divisible by all common divisors in G of the set $\mathfrak{a}$. The notion of an ideal can be extended to subsets $\mathfrak{a}$ of G without modifying the definition, i.e. $\mathfrak{a}$ is an ideal if it contains every element of G which is divisible by all common divisors of $\mathfrak{a}$, but we restrict ourselves to fractional ideals. If A is any fractional set, it is readily verified that $(A^{-1})^{-1}$ is precisely the set of all elements of G divisible by all common divisors of A, and is therefore the ideal generated by A. A subset $\mathfrak{a}$ of G is therefore an ideal if and only if $(\mathfrak{a}^{-1})^{-1} = \mathfrak{a}$.

Evidently $(A^{-1})^{-1} \supseteq A$, and if $A \supseteq B$ then $A^{-1} \subseteq B^{-1}$. From this it follows that A^{-1} is an ideal if A is any fractional set. If an ideal $\mathfrak{a}$ has the property that $\mathfrak{a}\mathfrak{a}^{-1} = (1)$ it is called *invertible*. The semigroup $\mathfrak{G}$ of all ideals in G is a group if and only if every ideal is invertible.

An element α of G is said to be *almost integral* with respect to S if there exists an element c in S such that all the elements $c\alpha^n$ ($n = 1, 2, 3, \cdots$) are in S. S is *completely integrally closed* if it contains all elements of G which are almost integral with respect to S.

LEMMA 5.1. *Every regular arithmetic S is completely integrally closed. If a regular semigroup S admits a regular ideal arithmetic Σ preserving the division relation in S (condition* $\mathbf{N}_1$*), then S is completely integrally closed.*

PROOF. Suppose that ab^{-1} is almost integral with respect to S. Then c exists in S such that the elements $c(ab^{-1})^n$ are in S, i.e. $b^n \mid ca^n$ $(n = 1, 2, \cdots)$. Since S is an arithmetic we can assume $(a, b) \sim 1$. Then $(a^n, b^n) \sim 1$ and $b^n \mid c$. But this is impossible unless b is a unit, whence ab^{-1} is in S.

To prove the second part we note that a regular semigroup S is completely integrally closed if and only if its reduced semigroup $\bar{S}$ has this property. Hence we may assume that S is itself reduced, and $S \subseteq \Sigma$. The quotient-group G of S may then be regarded as a subgroup of the quotient-group of Σ. Hence if ab^{-1} is an element of G almost integral with respect to S, it is almost integral with respect to Σ. But Σ is an arithmetic, and from the first part of the lemma it follows that ab^{-1} is in Σ. This means that $b \mid_\Sigma a$, and by $\mathbf{N}_1$ it follows that $b \mid_S a$, i.e. ab^{-1} is in S.

LEMMA[10] 5.2. *The semigroup $\mathfrak{G}$ of all fractional ideals in a regular semigroup S is a group if and only if S is completely integrally closed. Under these circumstances, if $\mathfrak{a} \supseteq \mathfrak{b}$ ($\mathfrak{a}$, $\mathfrak{b}$ in $\mathfrak{G}$) then $\mathfrak{a} \mid \mathfrak{b}$ relative to the semigroup $\mathfrak{S}$ of all integral ideals in S.*

PROOF. Assuming that S is completely integrally closed, we show that every ideal $\mathfrak{a}$ of $\mathfrak{G}$ is invertible. Let α be any element of the ideal $(\mathfrak{a}\mathfrak{a}^{-1})^{-1}$, i.e. any element of G such that $\alpha\mathfrak{a}\mathfrak{a}^{-1} \subseteq (1)$. Then $\alpha\mathfrak{a}^{-1} \subseteq \mathfrak{a}^{-1}$, whence $\alpha^n\mathfrak{a}^{-1} \subseteq \mathfrak{a}^{-1}$ $(n = 1, 2, \cdots)$ and $\alpha^n\mathfrak{a}\mathfrak{a}^{-1} \subseteq (1)$. Hence if c is any element of $\mathfrak{a}\mathfrak{a}^{-1}$ then $c\alpha^n$ is in S for all $n = 1, 2, \cdots$, i.e. α is almost integral with respect to S. α is therefore in S by hypothesis, which shows that $(\mathfrak{a}\mathfrak{a}^{-1})^{-1} \subseteq (1)$, whence $\mathfrak{a}\mathfrak{a}^{-1} \supseteq (1)$.

Assuming now that $\mathfrak{G}$ is a group, suppose that $\mathfrak{a} \supseteq \mathfrak{b}$. Then $(1) = \mathfrak{a}\mathfrak{a}^{-1} \supseteq \mathfrak{b}\mathfrak{a}^{-1}$, so that $\mathfrak{c} = \mathfrak{b}\mathfrak{a}^{-1}$ is an integral ideal. Since $\mathfrak{a}\mathfrak{c} = \mathfrak{b}\mathfrak{a}\mathfrak{a}^{-1} = \mathfrak{b}(1) = \mathfrak{b}$, it follows that $\mathfrak{a} \mid_{\mathfrak{S}} \mathfrak{b}$.

Let α be any element of G almost integral with respect to S. Then the ideal $(\alpha, \alpha^2, \alpha^3, \cdots)$ generated by all the powers of α is fractional, since there exists c in S such that all the elements $c\alpha^n$ $(n = 1, 2, \cdots)$ are in S. The same holds for $(1, \alpha, \alpha^2, \cdots)$. Since

$$(1, \alpha, \alpha^2, \cdots) \supseteq (\alpha, \alpha^2, \alpha^3, \cdots)$$

it follows from the above that there exists an integral ideal $\mathfrak{c}$ such that

$$(1, \alpha, \alpha^2, \cdots)\mathfrak{c} = (\alpha, \alpha^2, \alpha^3, \cdots) = (1, \alpha, \alpha^2, \cdots)(\alpha).$$

Since $\mathfrak{G}$ is a group we can cancel the ideal $(1, \alpha, \alpha^2, \cdots)$ on both sides. From $(\alpha) = \mathfrak{c}$ it follows that α is in S.

THEOREM 5.2. *For a regular semigroup S, the two conditions* **H** *and* **J′** *below are equivalent to the two conditions* **F** *and* **M**.

H. *If $\mathfrak{a} \supseteq \mathfrak{b}$ ($\mathfrak{a}$ and $\mathfrak{b}$ integral ideals in S), then $\mathfrak{a} \mid_{\mathfrak{S}} \mathfrak{b}$.*

J′. *The semigroup $\mathfrak{S}$ of integral ideals in S is a regular arithmetic.*

F. *Chain condition for finite integral ideals (Theorem 2.1)*

M. *S is completely integrally closed.*

---

[10] Krull, l.c., p. 120.

PROOF. Since $\mathbf{J}'$ is a stronger condition than $\mathbf{J}$ of Theorem 2.3, it follows from the latter that $\mathbf{H}$ and $\mathbf{J}'$ imply $\mathbf{F}$. From Theorem 4.1 we see that $\mathfrak{S}$ is a regular normal ideal arithmetic of S. $\mathbf{M}$ then follows from Lemma 5.1.

Conversely, $\mathbf{H}$ is a consequence of $\mathbf{M}$ by Lemma 5.2. $\mathbf{J}'$ is then proved by establishing $\mathbf{A}$ and $\mathbf{D}$ of Theorem 5.1 applied to $\mathfrak{S}$, as in the proof of Theorem 2.3.

THEOREM 5.3. *If a regular reduced semigroup S admits a regular ideal arithmetic Σ which preserves the division relation in S (condition $\mathbf{N}_1$), then the semigroup $\mathfrak{S}$ of integral ideals in S is a regular normal ideal arithmetic of S.*

PROOF. We proceed to show that S satisfies the conditions $\mathbf{F}$ and $\mathbf{M}$ of Theorem 5.2; then from $\mathbf{H}$ and $\mathbf{J}'$ it follows that $\mathfrak{S}$ is a regular normal ideal arithmetic of S (Theorem 4.1).

$\mathbf{M}$ follows from Lemma 5.1. To prove $\mathbf{F}$ we show first that if α is the GCD in Σ of an integral ideal $\mathfrak{a}$, then $\mathfrak{a}$ is precisely the set of all multiples of α in S. Since $\mathfrak{a}$ is evidently contained in this set, we need only show that any multiple c of α in S is in $\mathfrak{a}$. Suppose that s and t in S are such that $s \mid_S at$ for all a in $\mathfrak{a}$. Since αt is the GCD of the set of all at with a in $\mathfrak{a}$, it follows that $s \mid_\Sigma \alpha t$. Since $\alpha \mid_\Sigma c$ it follows that $s \mid_\Sigma ct$, and hence $s \mid_S ct$ by $\mathbf{N}_1$.

From this it follows that if $\mathfrak{a}$ and $\mathfrak{b}$ are integral ideals such that $\mathfrak{a} \supset \mathfrak{b}$, and if α and β are their respective GCD's in Σ, then $\alpha \mid\mid \beta$. The condition $\mathbf{F}$ is therefore a consequence of the divisor-chain condition $\mathbf{A}$ for elements of the arithmetic Σ.

From this theorem it follows incidentally that condition $\mathbf{H}$ of Theorem 5.2 is a consequence of $\mathbf{J}'$, for $\mathfrak{S}$ always satisfies $\mathbf{N}_1$. Hence if $\mathfrak{S}$ is a regular arithmetic, it must be a regular normal ideal arithmetic of S.

We conclude with a theorem which gives a (not altogether satisfactory) answer to the question of whether or not a regular semigroup admits a regular ideal arithmetic of any kind whatever.

THEOREM 5.4. *A regular semigroup S can admit a regular ideal arithmetic if and only if there exists a reduced semigroup S' between S and its quotient-group G, $S \subsetneq S' \subset G$, which admits a regular normal ideal arithmetic.*

PROOF. Let Σ be any regular ideal arithmetic of S, and let $S' = \Sigma \cap G$. Then Σ is a regular ideal arithmetic of S' preserving the division relation in S'. For suppose that $(ab^{-1})\xi = cd^{-1}$, with ab^{-1} and cd^{-1} in S' and ξ in Σ. Then $\xi = (bc)(ad)^{-1}$ is in both Σ and G, and hence in S'. By Theorem 5.3, S' admits a regular normal ideal arithmetic.

INSTITUTE FOR ADVANCED STUDY.

ANNALS OF MATHEMATICS
Vol. 39, No. 3, July, 1938

PERIODIC TRANSFORMATIONS OF COMPLEXES[1]

BY M. RICHARDSON AND P. A. SMITH

(Received June 1, 1937; Revised April 25, 1938)

1. Introduction

In this paper we shall consider certain simple topological properties of periodic homeomorphisms of a space S into itself. Two homeomorphisms T and T_0 of S into itself are said to belong to the same topological type if there is a third homeomorphism τ such that $T_0 = \tau T \tau^{-1}$. For arbitrary T's, many characteristic properties of types can be described in terms of the structure of certain invariant sets, the recurrence properties under repeated iteration and so on, (see for example [1]); but the problem of complete classification of types even for the case where S is, say, a sphere is too general to be of great interest. If, however, one considers only homeomorphisms of finite period, various algebraic type invariants can be defined. For orientable surfaces, for example, such invariants are actually sufficient to enable one to determine whether or not two sense-preserving periodic transformations belong to the same type (Nielsen [6]).

We shall here consider such invariants as can be readily defined in terms of homology theory. For the most part, we shall maintain a purely combinatorial point of view throughout and shall take S to be a finite simplicial complex and T to be a homeomorphism of S into itself which carries simplexes into simplexes, T being then necessarily of finite period. In the definition of "type," τ is also understood to be a simplicial homeomorphism. It will be seen that some of our considerations resemble those which have recently led Reidemeister and de Rham to the (strictly combinatorial) classification of lens- and cyclic-spaces. While our invariants do not have this sharpness, they do have the merit of being more topological in character; some of our theorems have meaning and hold true equally well for very general spaces. This implies in particular that when S is a complex, T and τ do not really need to be simplicial. Although the modifications necessary to place our results in a really topological setting will not be carried out here, it will readily be seen from our papers [7] and [8] what the required procedure would be.

Our invariants are closely related, as we shall show, to the homology groups of the so-called modular space of S—the space obtained by identifying points which are images of each other under powers of T. In the particular case in

[1] Some of the results of this paper were presented to the American Mathematical Society under this title by one of the authors on Sept. 9, 1937. Cf. Bull. Amer. Math. Soc. 43 (1937) p. 474.

which S is the totality of ordered sets of m points $x_1, \cdots, x_m$ of a given complex K and T is the transformation which carries each $(x_1, x_2, \cdots, x_m)$ into $(x_2, x_3, \cdots, x_1)$, the corresponding modular space is called the m-fold cyclic product of K. We shall determine completely the modulo m homology groups of cyclic products of any K for prime m.

2. Simplest invariants

Let T be a homeomorphism of a space S into itself of period m. We shall make frequent use of the expressions

$$\sigma = \sigma(T) = 1 + T + \cdots + T^{m-1},$$

$$\delta = \delta(T) = 1 - T.$$

They are to be thought of as operators on the homology classes of S relative to a given coefficient group $\mathfrak{g}$; in particular if S is a simplicial complex and if T merely permutes the simplexes of S, σ and δ may be thought of as operating equally well on the individual chains of S. *The only coefficient groups* $\mathfrak{g}$ *which we shall consider are* $\mathfrak{a}$, *the additive group of integers, and* $\mathfrak{m}$, *the integers reduced mod* m.

Suppose m is prime; then in the formal expansion of $(1 - T)^m$ all the coefficients except the two extreme ones are divisible by m. Hence $(1 - T)^m = 1 - T^m (\text{mod } m)$ and $(1 - T)^{m-1} = (1 - T^m)/(1 - T) = 1 + T + \cdots + T^{m-1}$ (mod m). Hence if $\mathfrak{g} = \mathfrak{m}$, m a prime, we have

$$\delta^m = 0 \text{ and } \delta^{m-1} = \sigma. \tag{2.1}$$

Suppose now that $S = K_n$, where K_n is a space whose homology groups relative to $\mathfrak{g}$ are defined in a topologically definite manner, and have finite bases for dimensions $\leqq n$ and vanish for dimensions $> n$. (K_n could be a finite n-complex, for example.) We can define simple type invariants of T in terms of homologies on K. Let $A_h^i(K; \mathfrak{g})$ be the subgroup of the homology group $H_h(K; \mathfrak{g})$ consisting of all the elements (homology classes) C such that $\delta^i C = 0$. The groups A_h^i are[2] type invariants. For suppose that τ is an arbitrary homeomorphism of K into itself. Then writing $\delta_0 = 1 - \tau T \tau^{-1}$, $C^0 = \tau C$, the relation $\delta^i C = 0$ implies that $\delta_0^i C^0 = 0$ and vice versa. For suppose $\delta^i C = 0$. Then

$$\begin{aligned} \delta_0^i C^0 &= (1 - \tau T \tau^{-1})^i \tau C \\ &= \tau (1 - T)^i C \\ &= \tau \delta^i C = 0. \end{aligned}$$

[2] We shall often omit parts of a symbol (in particular, dimension indices) when no confusion can arise.

Thus the correspondence $C \to C^0$ induces an isomorphism between the group A_h^i defined relative to T and the corresponding group defined relative to $\tau T \tau^{-1}$. Since $\delta^{i+1}C = 0$ when $\delta^i C = 0$, we have

$$A^1 \subset A^2 \subset \cdots . \tag{2.2}$$

Consider next the automorphisms induced by T and its powers in the group $H_h(K; \mathfrak{g})$. It is obvious that these automorphisms are also type invariants. Let $\Gamma_h^1, \cdots, \Gamma_h^{r(h)}$ be a minimal basis for H_h and suppose that[3]

$$T\Gamma_h^\alpha = g_{hj}^\alpha \Gamma_h^j . \tag{2.3}$$

Let $u_h = e - g_h$ where $g_h = \| g_{hj}^\alpha \|$ and e is the identity matrix. If the given basis for H_h is replaced by a second basis, the matrix $g = g_h$ in (2.3) will be replaced by cgc^{-1} where c is a matrix with elements in $\mathfrak{g}$ and where $\det c = \pm 1$ when $\mathfrak{g} = \mathfrak{a}$, and $\det c$ is prime to m when $\mathfrak{g} = \mathfrak{m}$. The matrix u is similarly transformed and so are the powers of g and u. Thus all invariants of g^α, u^β under transformations by matrices c are type invariants of T.

Such invariants are of course related to the groups A_h^i. Suppose for example that $\mathfrak{g} = \mathfrak{m}$, m a prime, or else that $\mathfrak{g} = \mathfrak{a}$ and that there are no h-dimensional zero-divisors. Then, the rank of A_h^i equals $r(h) - q_i$ where q_i is the rank[4] of the matrix u_h^i. For let C be an element of A_h^i, say $C = x_\alpha \Gamma_h^\alpha$. The condition $\delta^i C = 0$ satisfied by C is equivalent to the conditions

$$\sum_\alpha (e - g)_{\alpha\beta}^{(i)} x_\alpha = 0 \tag{2.4}$$

where $\| (e - g)_{\alpha\beta}^{(i)} \| = (e - g)^i = u^i$. A maximal set of independent elements for A_h^i is furnished by the cycles $x_\alpha \Gamma^\alpha$ corresponding to a maximal linearly independent set of solutions of (2.4), and the number of solutions in such a set is $r - q_i$.

Suppose now that $\mathfrak{g} = \mathfrak{m}$ and m is a prime. We shall show how, with the aid of the groups $A^i (= A_h^i)$ a homology basis for H_h can be chosen such that the matrix g_h assumes a fairly simple form. We observe first that $H_h(K; \mathfrak{m})$ is now the direct sum of $r(h)$ cyclic groups of order m, where $r(h)$ is now the h-dimensional Betti number of K, mod m.

From the relation $\delta^m = 0$ and the definition of A^i, it follows that $A^m = H$. If a group A is a direct sum of cyclic groups of prime order m then every subgroup of A is itself such a direct sum and is moreover a direct summand of A. It therefore follows from (2.2) that for $0 \leqq i \leqq m - 1$, A^i is a direct summand of A^{i+1}, and we may write $A^{i+1} = B^i + A^i$, $B^i \frown A^i = 0$. Since $A^0 = 0$, $B^0 = A^1$. Then

$$H = B^0 + B^1 + \cdots + B^{m-1}$$

[3] An index which, in a given term, occurs once as a superscript and once as a subscript, indicates summation over its allowable values.

[4] When $\mathfrak{g} = \mathfrak{m}$, the rank of a group (matrix) means throughout the maximal number of elements (rows) linearly independent mod m.

and the ranks τ_i of the groups B^i satisfy the relations

$$\tau_i = q_i - q_{i+1},$$
$$\tau_0 + \tau_1 + \cdots + \tau_{m-1} = r(h). \tag{2.5}$$

If X is an element of B^i then $\delta X \in B^{i-1}$. For, $\delta^{i+1}X = 0$ since $A^{i+1} \supset B^i$; hence $\delta^i \delta X = 0$ so that $\delta X \in A^i$. Now δX is not an element of A^{i-1}, for if it were, we would have $\delta^{i-1}\delta X = 0$, that is, $\delta^i X = 0$ which would imply that $X \in A^i$ whereas $A^i \frown B^i = 0$. Hence we conclude that $\delta X \in B^{i-1}$. Let Y^1, $\cdots$, Y^{τ_i} and X^1, $\cdots$ $X^{\tau_{i-1}}$ be independent bases for B^i and B^{i-1} respectively. Since $\delta Y^\alpha = Y^\alpha - TY^\alpha \in B^{i-1}$, we may write

$$TY^\alpha = a_\beta^{(i)\alpha} X^\beta + Y^\alpha. \tag{2.6}$$

It follows that *if a basis is chosen for each B^i, these bases taken together form a basis for H in terms of which the matrix g assumes the special form*

$$a = \begin{Vmatrix} e & 0 & 0 & \cdots & \cdot & \cdot \\ a^{(m-2)} & e & 0 & \cdots & \cdot & \cdot \\ 0 & a^{(m-3)} & e & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & & & e & 0 \\ \cdot & \cdot & & & a^{(0)} & e \end{Vmatrix}$$

where the block $a^{(i)}$ has τ_i rows and τ_{i-1} columns and the e's are identity blocks of the appropriate sizes. The matrix a must of course be of order m, that is, $a^m = e$. As a matter of fact it can be readily verified that every matrix of the same form as a and with coefficients in $\mathfrak{m}$, m a prime, is of order m.

The matrix a is not uniquely determined by g since the elements in the blocks $a^{(i)}$ depend on the choice of the bases for the B^i. Suppose that a new choice of these bases yields a second matrix a_0, with blocks $a_0^{(i)}$. It can be shown without difficulty that there exist matrices $b^{(i)}$, $c^{(i)}$ such that

$$a_0^{(i)} = b^{(i)} a^{(i)} c^{(i+1)} \qquad (i = 0, \cdots, m-2). \tag{2.7}$$

Hence the algebraic problem of finding invariants of g under similarities is equivalent to finding invariants, under the operation (2.7), of a system of matrices $a^{(0)}$, $\cdots$ $a^{(m-2)}$ when $a^{(i)}$ is of τ_i rows and τ_{i-1} columns. We shall, in this connection, do no more than make a remark concerning the numbers q_i.

We observe first that the rank of $a^{(i)}$ equals the number of rows of $a^{(i)}$, namely τ_i. For if the rank were less, there would exist relations $b_\alpha a_\beta^{(i)\alpha} = 0$ with b's not all zero. Then from (2.6), $b_\alpha \delta Y^\alpha = 0 = \delta b_\alpha Y^\alpha$ so that $b_\alpha Y^\alpha \in A^1 \subset A^i$ and therefore $b_\alpha Y^\alpha$ could not be in B^i which is absurd. It follows from this, since $a^{(i)}$ has τ_{i-1} columns, that $\tau_0 \geqq \tau_1 \geqq \cdots \geqq \tau_{m-1}$. Hence from (2.5), we have

$$2q_i \geqq q_{i-1} + q_{i+1} \qquad (i = 0, \cdots, m-1; q_{-1} = q_m = 0) \tag{2.8}$$

We have then the algebraic theorem that *if g is a square matrix with elements in* $\mathfrak{m}$ *and such that* $g^m = e$, *then the ranks* q_i *of the matrices* $(e - g)^i$ *must satisfy the relations* (2.8).

3. The special homology groups

We shall assume from now on that the complex K is simplicial and that the homeomorphism T merely permutes the simplexes of K, i.e. that T also is simplicial. We shall assume in addition that the points of K which are invariant under $T^h (1 \leqq h \leqq m - 1)$ constitute a subcomplex $L^{(h)}$ of K. In the sequel, $L^{(1)}$ will be denoted simply by L.

Let $\rho(T)$ be a polynomial in T with coefficients in $\mathfrak{g}$. We shall say that a chain X is a ρ-chain if $\rho X = 0$, and that X is a (ρ, L)-chain if $\rho X = 0 \bmod L$.

Suppose that H is an additive group of h-cycles relative or absolute, and that H' is a subgroup of H consisting of those cycles of H which bound in some special sense; then $H - H'$ can be considered as a "special" homology group. By the aid of ρ- and (ρ, L)-chains with various choices of $\mathfrak{g}$, one can define a variety of special homology groups[5] which are type invariants of T. For example, one may take for H the totality of ρ-cycles of dimension h and for H' those which are boundaries of ρ-chains; or one may take for H the totality of h-dimensional (ρ, L)-cycles mod L, and for H' those which are boundaries mod L of (ρ, L)-chains. The resulting difference groups, which we shall denote by H_h^ρ and $H_h^{\rho L}$ respectively, are, as here defined, only invariant in the combinatorial sense; that is they are invariant under transformations τ which are simplicial (see introduction). It will be seen however from [7] and [8] how they can be given purely topological meaning. H^ρ and $H^{\rho L}$ are by no means the only invariant groups which we can define in terms of the ρ- and (ρ, L)-chains, but they will illustrate various properties which groups of this sort have in general.

It will be convenient to assume from now on that all simplexes which are invariant under powers of T are also invariant under T. This of course holds automatically when m is prime. If m is composite, the assumption will be satisfied if T is replaced by T^k where k is a sufficiently large divisor of $m (k < m)$. Consequently we may regard the assumption not as a restriction on T but rather as a restriction on $\rho(T)$, since it is equivalent to the assertion that we shall consider only polynomials of the form $\rho(T^k)$ where k is a suitable divisor of m.

We shall limit our attention to the groups corresponding to the polynomials $\sigma(T)$, $\delta(T)$, $(\delta(T))^i$. We shall agree that the symbols ρ, $\bar{\rho}$ are to mean either σ or δ; when ρ is taken to mean $\sigma(\delta)$, then $\bar{\rho}$ is to mean $\delta(\sigma)$. When $\mathfrak{g} = \mathfrak{m}$ and m is prime, we shall allow ρ to stand also for δ^i and agree that $\bar{\rho}$ is then to mean δ^{m-i}. We note that in any case, $\rho\bar{\rho} = \bar{\rho}\rho = 0$ (see (2.1)).

---

[5] Cf. [8].

As a consequence of our assumptions concerning T, the h-simplexes of K can be denoted without repetition by

$$TE_h^i,\ T^2E_h^i,\ \cdots\ T^mE_h^i;\ E_h^{*j} \qquad (h = 0, \cdots, n)$$

where i and j have finite ranges and where the E^*'s are the simplexes of L. We shall employ a superscript * with the symbol X to signify that X is a chain in L, and a superscript 0 to signify that $X \frown L = 0$. Thus X^0 is a linear form in the T^kE^i, X^* a linear form in the E^{*j}. When $\mathfrak{g} = \mathfrak{m}$, every X^* is both a ρ- and a $\bar{\rho}$-chain.

3.1. *A necessary and sufficient condition that X be a (ρ, L)-chain is that there exists a chain Y^0 such that $X = \bar{\rho}Y^0$ mod L.*

Proof. That the condition is sufficient follows from the relation $\rho\bar{\rho} = 0$. We may write $X = b_{ai}T^aE^i$ mod L (dropping the dimensional subscript). Suppose first that X is a (δ, L)-chain. Then

$$\delta X = b_{ai}(T^a - T^{a+1})E^i = (b_{ai} - b_{a-1,i})T^aE^i = 0 \bmod L$$

the subscript a being taken mod m. Hence $b_{1i} = b_{2i} = \cdots = b_{mi}$ so that $X = \sigma Y^0$ mod L where $Y^0 = b_{1i}E^i$. Suppose next that X is a (σ, L)-chain. Then

$$\sigma X = \sum_{a=1}^{m} b_{ai}\sigma E^i = 0 \bmod L.$$

Hence $\sum_a b_{ai} = 0$ for all i. We wish to show the existence of a chain $Y^0 = x_{ai}T^aE^i$ such that $X = b_{ai}T^aE^i = \delta(x_{ai}T^aE^i) = (x_{ai} - x_{a+1,i})T^aE^i$. This amounts to showing the existence of solutions for the systems of equations

$$x_{1i} - x_{mi} = b_{1i},\ \cdots,\ x_{mi} - x_{m-1,i} = b_{mi}.$$

But on account of the relation $\sum_a b_{ai} = 0$, a solution for the i^{th} system is given by $x_{ai} = b_{1i} + b_{2i} + \cdots + b_{ai}$ $(a = 1, \cdots, m)$. Suppose finally that $\mathfrak{g} = \mathfrak{m}$, m a prime, and that $\rho = \delta^i$. We establish the necessity of our condition by induction on i; it has already been established for $i = 1$ since now $\delta^{m-1} = \sigma$. Suppose $\delta^iX = 0$ mod L. This may be written $\delta^{i-1}\delta X = 0$ mod L. By the hypothesis of the induction we may write $\delta X = \delta^{m-i+1}Y^0$ mod L. Let $D = X - \delta^{m-i}Y^0$. Then $\delta D = 0$ mod L, hence $D = \sigma W = \delta^{m-1}W$ mod L. Hence $X = \delta^{m-i}(Y^0 + \delta^{i-1}W)$ mod L.

3.2. *If $\mathfrak{g} = \mathfrak{m}$, every (ρ, L)-chain is a ρ-chain.*

Proof. Let $X = X^0 + X^*$ be a (ρ, L)-chain. Then $\rho X = \rho(X^0 + X^*) = \rho X^0 = 0$ mod L, which implies $\rho X^0 = 0$. Now since $TX^* = X^*$ we see that $\rho X^* = 0$ (modulo m) so that $\rho X = \rho X^0 = 0$.

3.3. *If $\mathfrak{g} = \mathfrak{m}$, a necessary and sufficient condition that X be a ρ-chain is that X be expressible in the form $\bar{\rho}Y^0 + Y^*$.*

Proof. If X is a ρ-chain then X is a (ρ, L)-chain and hence $X = \bar{\rho}Y^0$ mod L; hence $X = \bar{\rho}Y^0 + Y^*$. Conversely, if $X = \bar{\rho}Y^0 + Y^*$, then X is a (ρ, L)-chain, hence, by 3.2, a ρ-chain.

We now examine more closely the homologies defining the special homology groups. If C is a ρ-cycle and bounds a ρ-chain, we shall write $C \sim_\rho 0$; if C is a (ρ, L)-cycle mod L and bounds mod L a (ρ, L)-chain, we shall write $C \sim_\rho 0$ mod L. These special homologies obey the usual formal laws.

3.4. *If* $\mathfrak{g} = \mathfrak{m}$ *and if* $\rho C + C^*$ *is a cycle, then* ρC *and* C^* *are cycles. If in addition* $\rho C + C^* \sim_{\bar\rho} 0$, *then* $\rho C \sim_{\bar\rho} 0$ *and* $C^* \sim 0$ *on* L, *so that* $C^* \sim_{\bar\rho} 0$.

PROOF. If the boundary of C contains a simplex E^*, the simplex will occur in the boundary of ρC with the coefficient 0 since $\rho E^* = 0$. Consequently $F(\rho C)$ contains no simplex of L and hence $F(\rho C) = F(C^*) = 0$. Suppose that $\rho C + C^* \sim_{\bar\rho} 0$, say $\rho C + C^* = F(\rho D + D^*)$. Then by the same argument we conclude that $F(\rho D) \frown L = 0$, hence $\rho C = F(\rho D)$, $C^* = F(D^*)$.

3.5. *If* $C \sim 0$, *then* $\bar\rho C \sim_\rho 0$. *This holds equally well if cycles and homologies are taken mod* L.

PROOF (for the mod L case). Suppose that C is a cycle mod L and that $C = F(D)$ (say) mod L. Then $\bar\rho C = F(\bar\rho D) = \bar\rho F(D)$ mod L and the desired homology follows from the fact that $\bar\rho F(D)$ is a (ρ, L)-chain.

3.6. *If* $\bar\rho X$ *is a cycle mod* L *then* $F(X)$ *is a* $(\bar\rho, L)$*-cycle; and if* $\bar\rho X \sim_\rho 0$ *mod* L, *then* $F(X) \sim_{\bar\rho} 0$ *mod* L. *If* $\mathfrak{g} = \mathfrak{m}$, *the same statement holds with* L *and mod* L *omitted.*

PROOF. Since $F(\bar\rho X) = \bar\rho F(X) = 0$ mod L, it follows that $F(X)$ is a $(\bar\rho, L)$-cycle. The relation $\bar\rho X \sim_\rho 0$ mod L implies the existence of a chain Y such that $F(\bar\rho Y) = \bar\rho X$ mod L. Let $Z = X - F(Y)$. Then $\bar\rho Z = \bar\rho X - F(\bar\rho Y) = 0$ mod L so that Z is a $(\bar\rho, L)$-chain. Since $F(Z) = F(X)$ mod L, we have $F(X) \sim_{\bar\rho} 0$ mod L. If $\mathfrak{g} = \mathfrak{m}$ the same argument holds if L and mod L are everywhere omitted.

Let $\mathfrak{g} = \mathfrak{m}$ and let C be a ρ-cycle. We shall denote by $(C)_\rho$ the totality of ρ-cycles which are $\sim_\rho C$. The classes $(C)_\rho$ are precisely the elements of H^ρ. If we consider only classes $(C)_\rho$ where the ρ-cycle C is homologous to zero in the ordinary sense, we obtain a subgroup of H^ρ which we shall denote by B^ρ. By means of the preceding theorem, there can be defined a homomorphic mapping μ of H_h^ρ into $B_{h-1}^{\bar\rho}$. Let C_h and C_h' be ρ-cycles such that $C_h \sim_\rho C_h'$. We may write $C = \bar\rho X^0 + X^*$, $C' = \bar\rho X^{0\prime} + X^{*\prime}$. Then if C and C' belong to the same ρ-homology class of H^ρ, it follows from 3.4 and 3.6 that $F(X^0)$ and $F(X^{0\prime})$ belong to the same $\bar\rho$-homology class of $B_{h-1}^{\bar\rho}$. Thus there exists a single-valued correspondence μ of the form

$$(\bar\rho X_h^0 + X_h^*)_\rho \to (F(X_h^0))_{\bar\rho}$$

of elements of H_h^ρ into certain elements of $B_{h-1}^{\bar\rho}$. It is clear moreover that μ is additive and is therefore a homomorphism of H_h^ρ into a subgroup of $B_{h-1}^{\bar\rho}$. For arbitrary $\mathfrak{g}$, we define $B^{\rho L}$ as the group of (ρ, L)-homology classes of the form $(C)_{\rho L}$ where C is a (ρ, L)-cycle homologous to 0 mod L. It is easy to see that, in a similar way, there exists a homomorphism μ_L of $H_h^{\rho L}$ into $B_{h-1}^{\bar\rho L}$.

In order to determine the precise nature of the homomorphisms μ, μ_L, we introduce the subgroups K^ρ, $K^{\rho L}$ of H^ρ, $H^{\rho L}$, defined as follows: K^ρ consists of

those elements of H^ρ which are of the form $(\bar\rho C + C^*)_\rho$ where C and C^* are cycles; $K^{\rho L}$ consists of those elements of $H^{\rho L}$ which are of the form $(\bar\rho C)_{\rho L}$ where C is a cycle mod L.

3.7. THEOREM. *$H_h^{\rho L} - K_h^{\rho L} \cong B_{h-1}^{\bar\rho L}$. If $\mathfrak{g} = \mathfrak{m}$, then $H_h^\rho - K_h^\rho \cong B_{h-1}^{\bar\rho}$.*

PROOF. We need consider only the case $\mathfrak{g} = \mathfrak{m}$ since the proof for arbitrary $\mathfrak{g}$, mod L, is quite similar. We show first that the group into which μ maps H_h^ρ is exactly $B_{h-1}^{\bar\rho}$. Let $(C)_{\bar\rho}$ be an element of $B_{h-1}^{\bar\rho}$. We need to show that there exists an element of H_h^ρ which is carried by μ into $(C)_{\bar\rho}$. We may assume that there exists a chain Y_h such that $F(Y_h) = C$. Then $F(\bar\rho Y) = \bar\rho C = 0$ so that $\bar\rho Y$ is a ρ-cycle. Obviously the image under μ of $(\bar\rho Y)_\rho$ is precisely $(C)_{\bar\rho}$. It remains to show that K_h^ρ is the kernel of the mapping μ. If D_h and D_h^* are cycles, then μ carries $(\bar\rho D_h + D_h^*)_\rho$ into 0. Hence the kernel of μ contains K_h^ρ. An arbitrary element of the kernel is representable in the form $(\bar\rho A + A^*)_\rho$ where $F(A) \sim_{\bar\rho} 0$. This last relation implies the existence of a chain $Z_h^0 + Z_h^*$ such that $F(\rho Z_h^0 + Z_h^*) = F(A)$. Hence $\Gamma_h = A - \rho Z_h^0 - Z_h^*$ is a cycle and moreover $\bar\rho\Gamma_h = \bar\rho A$, so that $(\bar\rho A + A^*)_\rho$ is an element of K_h^ρ. This concludes the proof for the case $\mathfrak{g} = \mathfrak{m}$; for arbitrary $\mathfrak{g}$, obvious modifications yield the mod L result.

The groups $B^{\rho L}$, B^ρ consist of classes of cycles which are homologous to zero in the ordinary sense. We shall examine their remainder classes. For this purpose, consider the subgroup G^ρ of the (ordinary) homology group $H(K)$ of same dimension defined as the totality of[6] classes (C) such that C is a cycle satisfying the relation $\rho C \sim_{\bar\rho} 0$. For arbitrary $\mathfrak{g}$, let $G^{\rho L}$ be similarly defined.

3.8. THEOREM. *For each dimension, $H^{\rho L} - B^{\rho L} \cong G^{\rho L}$. When $\mathfrak{g} = \mathfrak{m}$, $H^\rho - B^\rho \cong G^\rho$.*

PROOF (for $\mathfrak{g} = \mathfrak{m}$). The correspondence $(X)_\rho \to (X)$, where X is a ρ-cycle, is a homomorphic mapping of H^ρ into a subgroup of $H(K)$ and its kernel is B^ρ. We shall show that the subgroup into which H^ρ is mapped is exactly G^ρ. Let $(C)_\rho$ be an arbitrary element of H^ρ, C being a ρ-cycle. Since $\rho C = 0 \sim_{\bar\rho} 0$, it follows that (C) is an element of G^ρ. Conversely, let (D) be an arbitrary element of G^ρ, D being a cycle such that $\rho D \sim_{\bar\rho} 0$. We assert that the class (D) contains a ρ-cycle. For there exists a chain Y^0 such[7] that $F(\rho Y^0) = \rho D$. Then if we put $\Delta = D - F(Y^0)$, the facts that Δ is a cycle and is $\sim D$, and that $\rho\Delta = 0$, prove that Δ is the desired ρ-cycle. Since $(\Delta) = (D)$, the mapping in question carries $(\Delta)_\rho$ into (D). This completes the proof for $\mathfrak{g} = \mathfrak{m}$; obvious modifications yield a proof for arbitrary $\mathfrak{g}$, mod L.

3.9. *The group $B^{\rho L}$ is finite and the order of each element is a divisor of m.*

PROOF. We need only show that each element x of $B^{\rho L}$ satisfies the equation $mx = 0$. Let $(C)_{\rho L}$ be an element of $B^{\rho L}$, C being a (ρ, L)-cycle mod L. We may assume the existence of a chain Y such that $F(Y) = C \bmod L$. We assert

---

[6] Hereafter, (X) shall denote the homology class of the cycle X.

[7] Since $\rho D \frown L = 0$ and $F\rho Y^0 \frown L = 0$, we see that $F(\rho Y^0 + Y^*) = \rho D \sim_{\bar\rho} 0$ implies $FY^* = 0$; hence $F\rho Y^0 = \rho D \sim 0$.

that $(1 - T)C \sim_\rho 0 \bmod L$. Suppose first that $\rho = \sigma$. Then $(1 - T)F(Y) = (1 - T)C \bmod L$ and our assertion follows from the fact that $(1 - T)Y$ is a (σ, L)-chain. If $\rho = \delta$, our assertion follows from the relation $(1 - T)C = 0 \sim_\delta 0 \bmod L$. We conclude that in either case

$$C \sim_\rho TC \sim_\rho T^2C \sim_\rho \cdots \sim_\rho T^{m-1}C \bmod L.$$

Thus if C is a (σ, L)-cycle mod L, we have

$$0 = \sigma C = (1 + T + \cdots + T^{m-1})C \sim_\sigma mC \bmod L.$$

Suppose that $\rho = \delta$. Then since C is a δ-cycle mod L, we may write $C = \sigma X$ mod L. Since $F(\sigma Y) = \sigma C = \sigma\sigma X = mC \bmod L$, we have $mC \sim_\delta 0 \bmod L$.

Let us assume that $\mathfrak{g}$ is the group of integers and that K possesses no coefficients of torsion mod L of dimensions $h, h + 1, \cdots, n$. Then from 3.9, it follows readily that $B_j^{\rho L}$ $(h \leqq j \leqq n)$ is identical with the subgroup of $H_j^{\rho L}$ consisting of all elements of finite order and is therefore a direct summand of $H_j^{\rho L}$. The relation 3.8 may now be written

$$H^{\rho L} \cong B^{\rho L} + G^{\rho L}.$$

Suppose m is prime. Then if $\tau(H)$ denotes the smallest number of elements which can constitute a basis for H, it is readily seen from the fact that $B_j^{\rho L}$ is now the direct sum of cyclic groups of order m, that[8]

$$\tau(H^{\rho L}) = \text{rank } B^{\rho L} + \text{rank } G^{\rho L} \tag{3.10}$$

and $G^{\rho L}$ contains no elements of finite order. Moreover, we have, from (3.7),

$$\text{rank } B_j^{\rho L} \leqq \tau(H_{j+1}^{\delta L}). \tag{3.11}$$

Let $A^{\rho L}$ be the group of (ordinary) homology classes mod L of the form (C) where C is a cycle mod L such that $\rho C \sim 0 \bmod L$. Let q^ρ be the rank of the matrix $e - g$ when $\rho = \delta$, and $e + g + g^2 + \cdots + g^{m-1}$ when $\rho = \sigma$ (see §2; the homology relations are now to be taken mod L). Then as in §2, rank $A_j^{\rho L} = R_j(K; L) - q_j^\rho$. But obviously $G_j^{\rho L}$ is a subgroup of $A_j^{\rho L}$ and hence

$$\text{rank } G_j^{\rho L} \leqq R_j(K; L) - q_j^\rho. \tag{3.12}$$

As a result of (3.10), (3.11), (3.12) we have

$$\tau(H_j^{\rho L}) \leqq \tau(H_{j+1}^{\delta L}) + R_j(K; L) - q_j^\rho$$

which yields, on giving j the values $h, h + 1, \cdots, n$:

$$\tau(H_j^{\rho L}) \leqq R_h(K; L) + \cdots + R_n(K; L) - q_h^\rho - q_{h+1}^\delta - \cdots - q_n^\nu \tag{3.13}$$

[8] By "rank $B^{\rho L}$" is meant the number of cyclic summands in the cited decomposition of $B^{\rho L}$. Rank $G^{\rho L}$ has the usual meaning (cf. footnote 4). We remark that $\tau(H)$ is the number of summands in the canonical decomposition of H into a direct sum of free cyclic groups and finite cyclic groups such that the order of each divides the order of the next.

where $\nu = \rho$ or $\bar{\rho}$. Had we taken $\mathfrak{g} = \mathfrak{m}$ and m a prime, then $\tau(H^p) = \text{rank } H^p$ and it can be immediately verified that the inequality just obtained still holds provided that the L is everywhere omitted. It is in this form that the inequality has been used elsewhere[9] by one of the present authors to obtain a generalization of Harnack's theorem concerning the maximum number of real branches which a real algebraic curve may have.

4. The modular space

Let k be the totality of non-ordered sets $\{x, Tx, \cdots, T^{m-1}x\}$ (x in K). If there is introduced into k a topology derived in a natural way from that of K (see [7]), the resulting space, frequently called the modular space of K, T, is obviously a type invariant of T. We shall point out certain ways in which the homology groups of k are related to the special homology groups of K.[10] These relations reveal themselves most simply when $\mathfrak{g} = \mathfrak{m}$ and it is this case alone that we shall consider.

The correspondence Λ defined by

$$\Lambda x = \{x, Tx, \cdots, T^{m-1}x\} \qquad (x \text{ in } K)$$

is obviously[11] a single-valued and continuous mapping of K into k. As a matter of fact, since T carries simplexes into simplexes, it is easy to see that k itself is a complex and that Λ carries simplexes of K into simplexes of k. In particular Λ transforms L homeomorphically into a subcomplex, of k which we shall denote by l. By imposing the condition of additivity, Λ can be thought of as operating on the chains of K carrying them into chains of like dimension in k. On account of its continuity, Λ preserves boundary relations, hence carries cycles into cycles and preserves relations of homology. We note in addition that $\Lambda\rho$, thought of as an operator on chains of K, is an annihilator; for since $\Lambda E = \Lambda TE$ for every simplex E of K, it follows that

$$\begin{aligned} \Lambda\delta E &= 0, \\ \Lambda\delta^i E &= 0, \\ \Lambda\sigma \mathrm{E} &= m\Lambda E = 0 \pmod{m}. \end{aligned} \tag{4.1}$$

If e_h is a simplex of k, there exists a simplex E_h of K such that $\Lambda E_h = \Lambda T^i E_h = e_h$. We shall denote the chain σE_h by $\Lambda' e_h$. By imposing the condition of additivity, the domain of operation of Λ' is extended to the totality of chains of k. It is easy to see that if c is an arbitrary chain of k, there exists a chain C in K such that $\Lambda' c = \sigma C$, $\Lambda C = c$. Hence $\Lambda' c$ is a δ-chain. Moreover, Λ' preserves boundary relations, hence cycles and homologies; in fact if c is a cycle in k and if $c \sim 0$ in k, then it is easy to see from the preceding remarks that $\Lambda' c \sim_3 0$ in K.

---

[9] Cf. a forthcoming article by P. A. Smith in the Bulletin of the Amer. Math. Soc.

[10] The Betti numbers of k over the integers and mod π^u, where π is a prime not dividing m, are related in a simple way to the matrices g of section 2. Cf. [4].

[11] For full details on the assertions in this paragraph, cf. [7] and [4].

4.2. *If $\bar{\rho}C \sim_\rho 0$, then ΛC is a cycle mod l and $\Lambda C \sim 0$ mod l in k; if, in addition, ΛC is a cycle, then $\Lambda C \sim$ a cycle of l, but, if ΛC is not a cycle, then $F\Lambda C \sim 0$ on l.*

PROOF. By hypothesis, there exists a chain Y such that $F(\bar{\rho}Y) = \bar{\rho}C$. Let $X = F(Y) - C$. Then $\bar{\rho}X = 0$ and hence we may write $X = \rho Z^0 + Z^*$. Since $F(Y) = C + \rho Z^0 + Z^*$, we have $F(\Lambda Y) = \Lambda C + \Lambda Z^* \sim 0$ in k, which proves the theorem.

Let $H_h^{*\rho}$ be the subgroup of H_h^ρ consisting of all the $\rho-$homology classes which contain cycles in L. It follows immediately from 3.4 that $H_h^{*\rho}$ *is isomorphic with* $H_h(L)$ (*the h-dimensional homology group of L*), *hence also with* $H_h(l)$, since l and L are homeomorphic. We assert in addition that

4.3. $H_h^\delta - H_h^{*\delta}$ *is isomorphic with* $H_h(k; l)$, *the homology group of k mod l.*

PROOF. Let $(\sigma X + X^*)_\delta$ be an arbitrary element of H^δ. Then by (4.1), ΛX is a cycle mod l in k. Let $(\Lambda X)_o$ be the class of cycles mod l which are $\sim \Lambda X$ mod l in k. Then by 4.2, the correspondence $(\sigma X + X^*)_\delta \to (\Lambda X)_o$ is a homomorphic mapping of H^δ into a subgroup of $H(k; l)$. Let $(c)_o$ be the totality of cycles mod l in k which are $\sim c$ mod l, (c itself being such a cycle). Then $(c)_o$ may be considered as an arbitrary element of $H(k; l)$. From the remarks above, we may write $\Lambda' c = \sigma C$, $\Lambda C = c$. The mapping in question carries $(\Lambda' c)_\delta = (\sigma C)_\delta$ into $(\Lambda C)_o = (c)_o$ so that the image of H^δ is exactly $H(k; l)$. Now clearly $H^{*\delta}$ is contained in the kernel of the mapping. If $(\sigma X + X^*)_\delta$ is an arbitrary element of the kernel, which implies that $\Lambda X \sim 0$ mod l, then $\Lambda'\Lambda X = \sigma X \sim_\delta 0$. Therefore $(\sigma X + X^*)_\delta$ is the same as $(X^*)_\delta$ and is therefore an element in $H^{*\delta}$. Hence the kernel is precisely $H^{*\delta}$.

It will now be convenient to introduce a correspondence μ^* which resembles the mapping μ considered in §3. Let $(\bar{\rho}X_h + X_h^*)_\rho$ be an arbitrary element of H_h^ρ. The image under μ of this element is the element $(FX_h)_{\bar{\rho}}$. Since FX_h is a $\bar{\rho}$-cycle, we may write $FX_h = \rho X_{h-1}^0 + X_{h-1}^*$. Now if FX_h is replaced by any other representative of $(FX_h)_{\bar{\rho}}$, say $\rho Z_{h-1}^0 + Z_{h-1}^*$, so that $(\rho X_{h-1}^0 + X_{h-1}^*)_{\bar{\rho}} = (\rho Z_{h-1}^0 + Z_{h-1}^*)_{\bar{\rho}}$, then from 3.4, $X_{h-1}^* \sim Z_{h-1}^*$ in L. It follows therefore that the correspondence μ^* given by

$$(\bar{\rho}X_h^0 + X_h^*)_\rho \to (\Lambda X_{h-1}^*)_*,$$

where $(\Lambda X_{h-1}^*)_*$ means the class of cycles in l homologous in l to ΛX_{h-1}^*, determines a homomorphic mapping of H_h^ρ into a subgroup of $H_{h-1}(l)$. Since Λ defines a homeomorphism between L and l, it maps $H_h(L)$ isomorphically on $H_h(l)$. Let $\theta_h(l)$ be the subgroup of $H_h(l)$ consisting of those homology classes whose cycles are ~ 0 in k.

4.4. THEOREM. *The image under μ^* of H_h^δ is $\theta_{h-1}(l)$.*

PROOF. Let $(\sigma X_h + X_h^*)_\delta$ be an arbitrary element of H_h^δ. Then the element which corresponds to it under μ^* is $(\Lambda X_{h-1}^*)_*$ where $FX_h = \delta X_{h-1}^0 + X_{h-1}^{*\prime}$. Now $\Lambda FX_h = F\Lambda X_h = \Lambda X_{h-1}^*$ since $\Lambda\delta = 0$; therefore $\Lambda X_{h-1}^* \sim 0$ in k, and therefore is an element of $\theta_{h-1}(l)$. Conversely, suppose that $(\Lambda C_{h-1}^*)_*$ is an element in $\theta_{h-1}(l)$. This implies that $\Lambda C_{h-1}^* \sim 0$ in k so that there exists a chain d_h in k, bounded by ΛC_{h-1}^*. Then $\Lambda' d_h$ ($= D_h$, say) is bounded by $\Lambda'\Lambda C_{h-1}^* =$

$mC^*_{h-1} = 0 \pmod m$. That is D_h is a δ-cycle. We shall show that under μ^*, the image of $(D_h)_\delta$ is $(\Lambda C^*_{h-1})_*$. We may write $D_h = \sigma A_h$, where $\Lambda A_h = d_h$. Since σA_h is a cycle, we have $\sigma F A_h = 0$ and therefore $FA_h = \delta Y^0 + Y^*$ say. Therefore $Fd_h = F\Lambda A_h = \Lambda Y^*$ so that $\Lambda Y^* \sim 0$ in k. Consequently $(\Lambda Y^*_{h-1})_*$ is an element of $\theta_{h-1}(l)$ and this element is the image of $(D_h)_\delta$ under μ^*. But since $Fd_h = \Lambda C^*_{h-1}$, we have $\Lambda Y^*_{h-1} = \Lambda C^*_{h-1}$, hence $(\Lambda Y^*_{h-1})_* = (\Lambda C^*_{h-1})_*$ and the proof is complete.

The kernel Φ^δ_h of the mapping μ^* is the subgroup of H^δ_h composed of all δ-homology classes of the form $(\sigma X^0 + X^*)_\delta$ such that the part of FX^0 which is in L is a cycle homologous to zero in L. Thus,

$$\theta_{h-1}(l) \cong H^\delta_h - \Phi^\delta_h. \tag{4.5}$$

The groups Φ^δ_h furnish a link between the special homology groups of K and the ordinary homology groups of k, at least when $\mathfrak{g} = \mathfrak{m}$. This can be stated in a precise manner when m is prime, for then we can appeal to the following duality formula of Lefschetz:[12]

$$\mathrm{r}\, H_h(k) = \mathrm{r}\, H_h(k; l) + \mathrm{r}\, H_h(l) - \mathrm{r}\, \theta_h(l) - \mathrm{r}\, \theta_{h-1}(l). \tag{4.6}$$

Here r stands for "the rank of". Recalling that for prime m and $\mathfrak{g} = \mathfrak{m}$ every subgroup of a homology group (ordinary or special) is a direct summand, it follows from (4.5) that $\mathrm{r}\theta_{h-1}(l) = \mathrm{r}H^\delta_h - \mathrm{r}\Phi^\delta_h$. Moreover, from 4.3, $\mathrm{r}H_h(k; l) = \mathrm{r}H^\delta_h - \mathrm{r}H^{*\delta}_h = \mathrm{r}H^\delta_h - \mathrm{r}H_h(l)$. Combining these relations with (4.6), we have

4.7. THEOREM. *For prime m and $\mathfrak{g} = \mathfrak{m}$, the Betti numbers of the modular complex k are given by*

$$R_{h-1}(k) = \mathrm{r}\, \Phi^\delta_{h-1} + \mathrm{r}\, \Phi^\delta_h - \mathrm{r}\, H^\delta_h.$$

5. Cyclic Products

When m is composite, the determination of the homology characters of the modular space k seems to be a less simple matter than for prime m; for in order to obtain relations such as 4.7 it would first be necessary to find what form the duality relation (4.6) will take when m is composite—a problem somewhat complicated by the presence of "divisors of zero." In certain interesting cases however the characters of k can be obtained without appealing explicitly to any duality formula. Such is the case for example when k is a cyclic product of prime order (see introduction) and in the generalization of section 6. We shall study the topology of cyclic products in some detail, partly because we can in this way exhibit concrete examples of the various special groups, and partly because the results obtained appear to have interesting connections with certain problems in dynamics and the calculus of variations.

We continue to take $\mathfrak{g} = \mathfrak{m}$. Let K_n be a complex having no coefficients of torsion, i.e. a complex whose homology groups are direct sums of cyclic groups

[12] Cf. [2] p. 150.

of order m; every K_n has this property when m is prime. Let P_{mn} be the m-fold product

$$K_n \times K_n \times \cdots \times K_n$$

where the m factors are distinct copies of K_n. Let T be the homeomorphism which carries an arbitrary point $a \times b \times \cdots \times c$ of P into $b \times \cdots \times c \times a$ (factors permuted cyclically). T is of period m and *the modular space p_{mn} of P_{mn} relative to T is called the m-fold cyclic product of K.* We shall determine the special homology groups of P and the ordinary homology groups of p.

From the elementary properties of the products of complexes it is easy to deduce the following facts.[13] (1) There exists a simplicial subdivision of P, to be denoted also by P, such that the conditions in §3 are satisfied when m is prime; in particular the invariant points of T constitute a subcomplex L of P, and L is homeomorphic to K. (2) If C_h^j ($j = 1, \cdots R_h(K)$) are the cycles of a minimal homology basis for $H_h(K)$, then a minimal basis for $H_h(P)$ is given by the cycles

$$C_{h_1}^{j_1} \times C_{h_2}^{j_2} \times \cdots \times C_{h_m}^{j_m} \tag{5.1}$$

where $\sum h_i = h$, the factors being cycles in the respective copies of K. (3) Every cycle of the form

$$\Delta_{mq} = C_q \times C_q \times \cdots \times C_q$$

is transformed by T into itself or into its negative according to the rule[14]

$$T\Delta_{mq} = (-1)^e \Delta_{mq}, \qquad e = q(mq - 1) \bmod 2. \tag{5.2}$$

The cycles (5.1) which are not of this type fall when m is prime into sets such that each set contains m distinct cycles of equal dimension which are permuted cyclically among themselves by T. Thus for each dimension h, there exists a minimal base for $H_h(P)$ of the form

$$T^i \Gamma_h^j, \Delta_h^l \qquad (i = 0, \cdots, m - 1)$$

where the ranges for j, l are say $j = 1, \cdots, R_h^\Gamma$ and $l = 1, \cdots, R_h^\Delta$ so that $R_h(P)$ (the Betti number of P) equals $mR_h^\Gamma + R_h^\Delta$. The numbers R_h^Γ and R_h^Δ can of course be computed easily in terms of the Betti numbers of K. In particular $R_{mq}^\Delta = R_q(K)$ whereas $R_h^\Delta = 0$ if h is not a multiple of m.

In what follows, we shall assume that m is prime. Since $q(mq - 1)$ is then even, it follows from (5.2) that each cycle Δ_{mq} is transformed into itself. Thus each Δ_{mq} is a δ-cycle and, since $\mathfrak{g} = \mathfrak{m}$, it is a σ-cycle as well. When m is prime, each Δ_{mq} is equally well a δ^i-cycle. To indicate the fact that Δ_{mq} is to be regarded as a ρ-cycle[15] in a given discussion, we affix a left superscript ρ: ${}^\rho\Delta_{mq}$.

---

[13] Cf. [4] pp. 54–56.

[14] Cf. [4] p. 61.

[15] Hereafter, let $\rho = \sigma, \delta$ only.

5.3. *Let C_h be a ρ-cycle. Then $C_h \sim$ a linear combination of the cycles $\bar{\rho}T^i\Gamma_h^j$, Δ_h^l.*

PROOF. There is no essential restriction if we assume that there exists at most one Γ_h and at most one Δ_h. Then we may write $C_h \sim x_i T^i\Gamma_h + y\Delta_h$. Since $\rho\Delta = \rho C = 0$, we have $x_i\rho T^i\Gamma_h \sim 0$. According as $\rho = \delta$ or $\rho = \sigma$, the last relation may be written

$$\sum_{i=0}^{m-1}(x_i - x_{i-1})T^i\Gamma_h \sim 0 \text{ where } x_{-1} = x_{m-1}, \quad \text{or } \left(\sum_{i=0}^{m-1} x_i\right)\sigma\Gamma_h \sim 0.$$

Since the cycles $T^i\Gamma_h$ are part of a minimal basis, we conclude that $x_0 = x_1 = \cdots = x_{m-1} = x$ (say) or that $\sum x_i = 0$ as the case may be. In the first case therefore we have $C_h \sim x\sigma\Gamma_h + y\Delta_h$ and in the second we may write $C_h \sim \sum_{i=0}^{m-1} y_i\delta T^i\Gamma_h + y\Delta_h$ where $y_i = x_0 + x_1 + \cdots + x_i$, proving the theorem.

Let $h = mq$ and consider the cycle

$$^{\rho}\Delta_h = C_q \times \cdots \times C_q.$$

We may write $^{\rho}\Delta_h = {}^{\rho}\Delta_h^* + \bar{\rho}X_h$ where by 3.4, both terms in the sum are cycles. From the relation $0 = F\bar{\rho}X_h = \bar{\rho}FX_h$, it follows that FX_h is a $\bar{\rho}$-cycle; call it $^{\rho}\Delta_{h,h-1}$. The second subscript indicates the dimension, while the index ρ will be understood to refer to the character of the original Δ_h rather than $\Delta_{h,h-1}$. We now write $^{\rho}\Delta_{h,h-1} = {}^{\rho}\Delta_{h,h-1}^* + \rho^{\rho}X_{h,h-1}$ and denote the ρ-cycle $F^{\rho}X_{h,h-1}$ by $^{\rho}\Delta_{h,h-2}$. We continue in this manner down to the dimension 0. We assert that

$$^{\rho}\Delta_{mq}^* = {}^{\rho}\Delta_{mq,mq-1}^* = \cdots = {}^{\rho}\Delta_{mq,q+1}^* = 0. \tag{5.4}$$

For, T induces a homeomorphism of the complex $\bar{\Delta}_{mq}$ (closure of Δ_{mq}) into itself which is identical with the transformation which would be obtained if $\bar{C}_q$ were the basic complex instead of K_n and $\bar{\Delta}_{mq}$ played the part of P_{mn}. Hence the points of $\bar{\Delta}_{mq}$ which are invariant under T form a subcomplex of q dimensions and the truth of (5.4) is then obvious from the fact that the successive cycles Δ_{mq}, $\Delta_{mq,mq-1}$, $\cdots$ are all contained in $\bar{\Delta}_{mq}$. To summarize: *each $^{\rho}\Delta_{mq}^i$ gives rise*[16] *to a sequence*

$$^{\rho}\Delta_{mq,mq}^i, \ {}^{\rho}\Delta_{mq,mq-1}^i, \ \cdots \qquad ({}^{\rho}\Delta_{mq,mq}^i = {}^{\rho}\Delta_{mq}^i)$$

where, for $mq \geqq k \geqq q+1$

$$^{\rho}\Delta_{mq,k}^i = \begin{cases} \bar{\rho}^{\rho}X_{mq,k}^i = \text{a } \rho\text{-cycle when } mq\text{-}k \text{ is even}, \\ \rho^{\rho}X_{mq,k}^i = \text{a } \bar{\rho}\text{-cycle when } mq\text{-}k \text{ is odd}, \end{cases} \tag{5.5}$$

$$F^{\rho}X_{mq,k}^i = {}^{\rho}\Delta_{mq,k-1}^i.$$

[16] Cf. the homomorphism μ of section 3.

5.6. THEOREM I_h. *The following ρ-cycles constitute a basis for h-dimensional ρ-cycles relative to ρ-homologies:*

$$\bar{\rho}T^i\Gamma_h^j \qquad (j = 1, \cdots, R_h^\Gamma) \tag{5.7}$$

$$^\sigma\Delta_{ms,h}^b \qquad (b = 1, \cdots, R_{ms}^\Delta) \tag{5.8}$$

$$^\delta\Delta_{mt,h}^c \qquad (c = 1, \cdots, R_{mt}^\Delta) \tag{5.9}$$

where $i = 0$ when $\rho = \delta$ and $i = 0, \cdots, m-2$ when $\rho = \sigma$; s takes on all integer values such that $ms \geqq h \geqq s$, $ms - h$ even (odd) when $\rho = \sigma(\delta)$; similarly $mt \geqq h \geqq t$, $mt - h$ odd (even) if $\rho = \sigma(\delta)$. II_h. The cycles $^\sigma\Delta_{mh,h}^{*b}$ are linearly independent with respect to ordinary homologies in L, and constitute a basis for $H_h(L)$; the same is true of the cycles $^\delta\Delta_{mh,h}^{*c}$.

PROOF (by induction on h). Suppose $h = mn$. Since L is of dimension n, II_{mn} is trivial; and since the relations $\sim$, $=$, $\sim_\rho$ are now equivalent as far as ρ-cycles are concerned, we conclude from 5.3, that a given ρ-cycle C_{mn} is ρ-homologous to a linear combination of the cycles $^\rho\Delta_{mn}^d$ and the cycles $\bar{\rho}T^i\Gamma_{mn}^j$ $(i = 0, \cdots, m-1)$. Consider a sum $S = x_{ij}\bar{\rho}T^i\Gamma^j$. If $\bar{\rho} = \sigma$, then $S = \sum_i x_{ij}\sigma\Gamma^j$ which can be written in the form $g_{oj}\sigma T^0\Gamma^j$. If $\bar{\rho} = \delta$, S equals a linear combination of the cycles $\delta T^0\Gamma^j, \cdots, \delta T^{m-2}\Gamma^j$, since

$$\delta T^0\Gamma^j + \delta T^1\Gamma^j + \cdots + \delta T^{m-1}\Gamma^j = \delta\sigma\Gamma^j = 0.$$

In either case, therefore, we may assume that the index i in S equals 0 when $\rho = \delta$ and that $i = 0, \cdots, m-2$ when $\rho = \sigma$. This concludes the proof of I_{mn}.

Suppose now that I_i and II_i are true for $h < i \leqq mn$. Let C_h be a ρ-cycle. Then by 5.3, there exists a relation of the form

$$F(Y_{h+1}) = C_h + x_{ij}\bar{\rho}T^i\Gamma^j + y_d\,{}^\rho\Delta_{h,h}^d \sim 0, \tag{5.10}$$

and we may assume, as we have just shown, that i ranges over the values stated in the theorem. Since $F\rho Y = \rho FY = 0$, ρY is a $\bar{\rho}$-cycle and by the hypothesis of the induction we may write

$$\rho Y + z_{ij}\rho T^i\Gamma_{h+1}^j + v_k^s\,{}^\sigma\Delta_{ms,h+1}^k + w_l^t\,{}^\delta\Delta_{mt,h+1}^l \sim_{\bar{\rho}} 0, \tag{5.11}$$

where $s \geqq h+1$, $t \geqq h+1$ and $ms - (h+1)$ is even and $mt - (h+1)$ odd if $\bar{\rho} = \sigma$, vice versa if $\bar{\rho} = \delta$. The cycles in the first two terms contain no cells of L, and in the last two terms, the only cycles which can contain cells of L are those with $s = h+1$, $t = h+1$. Since $ms - (h+1)$ and $mt - (h+1)$ are of opposite parity, only one of the last two equalities can hold. Suppose it is the former; that is, suppose that s may assume the value $h+1$ whereas $t > h+1$. Then (5.11) can be written in the form

$$\rho Z_{h+1}^0 + v_l^s\,{}^\sigma\Delta_{ms,h+1}^{*l} \sim_{\bar{\rho}} 0 \qquad (s = h+1 \text{ not summed}).$$

By 3.4, the second term is ~ 0 in L and hence by II_{h+1} we conclude that the coefficients v_l^{h+1} are all 0. It follows that no cycle in (5.11) can contain cells of L and consequently (5.11) may be written in the form

$$\rho(Y + z_{ij}T^i\Gamma^j_{h+1} + v_k^s\,{}^\sigma X^k_{ms,h+1} + w_k^t\,{}^\delta X^k_{mt,h+1}) \sim_{\tilde\rho} 0.$$

On application of 3.6, we obtain $F(Y + \cdots) \sim_\rho 0$ and on replacing FY by the right side of (5.10), and the FX's by the cycles given in (5.5), we obtain a relation which establishes I_h.

It remains to prove II_h. Let D_h^* be a cycle in L. We may regard D_h^* as a σ-cycle and hence by I_h, $D^* \sim_\sigma$ a linear combination of the σ-cycles (5.7), (5.8), (5.9). Of these cycles, the only ones which contain cells of L are the ${}^\sigma\Delta^i_{mh,h}$. (Note that since m is assumed odd, $mh - h$ is even and hence ${}^\delta\Delta^i_{mh,h}$ can not be a σ-cycle.) Thus we may write

$$D^* \sim_\sigma u_i\,{}^\sigma\Delta^{*i}_{mh,h} + Z^0.$$

Hence by 3.4,

$$D^* \sim u_i\,{}^\sigma\Delta^{*i}_{mh,h} \text{ in } L.$$

Since D^* was arbitrary in L, it follows that the homology[17] group generated by the cycles ${}^\sigma\Delta^{*i}_{mh,h}$ is precisely $H_h(L)$. The number of these Δ^*'s is $R^\Delta_{mh} = R_h(K) = R_h(L)$ so that the basis which they constitute is minimal and they are therefore independent relative to homologies in L. Had we regarded D^* as a δ-cycle we would have obtained the same result for the cycles ${}^\delta\Delta^{*i}_{mh,h}$.

5.12. *The ρ-cycles of theorem* 5.6 *are independent with respect to ρ-homologies.*

PROOF. This is readily verified for the dimension mn. Assume that the theorem is true for the dimensions $mn, \cdots, h+1$. Let us suppose for definiteness that $\rho = \sigma$. Suppose there exists a σ-homology of the form

$$(5.13)\qquad F(\delta Z^0 + Z^*) = x_{ij}\delta T^i\Gamma^j + u_a\,{}^\sigma\Delta^a_h + u_b^s\,{}^\sigma\Delta^b_{ms,h} + v_c^t\,{}^\delta\Delta^c_{mt,h} \sim_\sigma 0$$

where $i = 0, \cdots, m-2$, and where $ms - h$ is even, $mt - h$ odd so that all the Δ's are σ-cycles. The cycles ${}^\sigma\Delta^a_h (= {}^\sigma\Delta^a_{h,h})$ exist only when h is a multiple of m, and since these cycles appear in a term by themselves, we may assume that ms does not assume the value h. Of course $mt \neq h$ because ${}^\delta\Delta_{h,h}$ could only represent a δ-cycle. It follows that the cycles in the last two terms are all boundaries of chains ${}^\sigma X_{ms,h+1}$, ${}^\delta X_{mt,h+1}$ (see (5.5)) hence are ~ 0. Now it is easy to verify that the cycles $\delta\Gamma^j, \cdots, \delta T^{m-2}\Gamma^j$ are independent with respect to ordinary homologies and of course they are also independent of the Δ^a_h's. It follows that the coefficients x_{ij} and u_a in (5.13) must all be 0. Thus (5.13) becomes

$$\begin{aligned} F(\delta Z^0 + Z^*) &= u_b^s\,{}^\sigma\Delta^b_{ms,h} + v_c^t\,{}^\delta\Delta^c_{mt,h} \\ &= F(u_b^s\,{}^\sigma X^b_{ms,h+1} + v_c^t\,{}^\delta X^c_{mt,h+1}) \\ &= FY, \text{ say.} \end{aligned}$$

[17] Relative to L.

Since $Y - \delta Z^0 - Z^*$ is a cycle, we may write

$$Y - \delta Z^0 - Z^* \sim \alpha_{ij} T^i \Gamma^j_{k+1} + \beta_k \Delta^k_{k+1}$$

and, on applying σ to both sides, we obtain by 3.5,

$$\sigma Y \sim_\delta \alpha_{ij} T^i \sigma \Gamma^j_{k+1}. \tag{5.14}$$

The index i here ranges from 0 to $m - 1$. But since $T^i\sigma = \sigma$, the right side of (5.14) may be written in the form $\alpha_j \sigma \Gamma^j_{k+1}$; that is, we may assume that the index i assumes only the value 0. Now

$$\sigma Y = u^s_b {}^\sigma\Delta^b_{mk,k+1} + v^t_c \Delta^c_{mt,k+1},$$

all cycles here being δ-cycles. Consequently (5.14) is a δ-homology among the δ-cycles of dimension $k + 1$ which are supposed, by the hypothesis of the induction, to be independent. Hence the u's and v's are 0, and this proves the theorem for the dimension k with $\rho = \sigma$. The modifications necessary for $\rho = \delta$ are obvious.

5.15. $F\Lambda\, {}^\rho X_{mk,k+1} = \Lambda\, {}^\rho\Delta^*_{mk,k} \sim 0$ in p.

Proof. $F\, {}^\rho X_{mk,k+1} = {}^\rho\Delta_{mk,k} = \lambda\, {}^\rho X_{mk,k} + {}^\rho\Delta^*_{mk,k}$ (where $\lambda = \rho$ or $\bar\rho$) and the desired relations follow from the fact that $\Lambda\lambda = 0$ when $\mathfrak{g} = \mathfrak{m}$.

5.16. *Corollary. Every cycle in l bounds in p.* For since the cycles ${}^\sigma\Delta^{*i}_{mk,k}$ generate the group $H_k(L)$ (Cf. 5.6, II_k), the cycles $\Lambda\, {}^\sigma\Delta^{*i}_{mk,k}$ generate $H_k(l)$.

5.17. *For $k \geqq t + 2$, $\Lambda\, {}^\rho X_{mt,k}$ is a cycle in p (to be denoted by ${}^\rho\xi_{mt,k}$).*

Proof. $F\Lambda X_{mt,k} = \Lambda F X_{mt,k} = \Lambda\Delta_{mt,k-1}$. Since $k - 1 > t$, we have $\Delta^*_{mt,k-1} = 0$. Hence $\Delta_{mt,k-1} = \rho X_{mt,k-1}$ so that $\Lambda\Delta_{mt,k-1} = 0$.

5.18. Theorem. *Every cycle of dimension k in p is homologous to a linear combination of the cycles ${}^\sigma\xi^b_{ms,k}$, ${}^\delta\xi^c_{mt,k}$, $\Lambda\Gamma^j_k$ where $s + 2 \leqq k$, $ms - k$ odd; $t + 2 \leqq k$, $mt - k$ even.*

Proof. Let γ_k be a cycle in p. Then $\Lambda'\gamma_k$ is a δ-cycle in P and we may write $\Lambda'\gamma_k = \sigma C_k$, $\Lambda C_k = \gamma_k$. We have by theorem 5.6,

$$\sigma C_k \sim_\delta x_i \sigma \Gamma^i_k + u^s_b {}^\sigma\Delta^b_{ms,k} + v^t_c {}^\delta\Delta^c_{mt,k}, \tag{5.19}$$

where $ms - k$ is odd, $mt - k$ even, all cycles now being δ-cycles. Since m is odd, $mk - k$ is even and therefore $s \neq k$, and hence the Δ^* of every ${}^\sigma\Delta^b_{ms,k}$ which can appear in (5.19) must be 0. Moreover, no ${}^\delta\Delta^{*c}_{mk,k}$ can appear with non-zero coefficient; for otherwise (5.19) would imply that some linear combination of the cycles ${}^\delta\Delta^{*c}_{mk,k}$ is ~ 0 on L, by 3.4, contradicting theorem 5.6, II_k . Hence every $v^k_c = 0$. Since $mk - (k - 1)$ is odd (m being odd) it follows that $l \neq k - 1$. Hence we may assume that t only ranges over values $\leqq k - 2$. Suppose now that some coefficient u^{k-1}_b is $\neq 0$. Then (5.19) can be written

$$0 \sim_\delta \sigma(C_k + x_i \Gamma^i + u^{k-1}_b X^b_{m(k-1),k} + \cdots) = \sigma Y$$

By 3.4, the second term is ~ 0 in L and hence by II_{h+1} we conclude that the coefficients v_l^{h+1} are all 0. It follows that no cycle in (5.11) can contain cells of L and consequently (5.11) may be written in the form

$$\rho(Y + z_{ij}T^i\Gamma^j_{h+1} + v^s_k\,{}^{\sigma}X^k_{ms,h+1} + w^t_k\,{}^{\delta}X^k_{mt,h+1}) \sim_{\bar{\rho}} 0.$$

On application of 3.6, we obtain $F(Y + \cdots) \sim_{\rho} 0$ and on replacing FY by the right side of (5.10), and the FX's by the cycles given in (5.5), we obtain a relation which establishes I_h.

It remains to prove II_h. Let D^*_h be a cycle in L. We may regard D^*_h as a σ-cycle and hence by I_h, $D^* \sim_{\sigma}$ a linear combination of the σ-cycles (5.7), (5.8), (5.9). Of these cycles, the only ones which contain cells of L are the ${}^{\sigma}\Delta^i_{mh,h}$. (Note that since m is assumed odd, $mh - h$ is even and hence ${}^{\delta}\Delta^i_{mh,h}$ can not be a σ-cycle.) Thus we may write

$$D^* \sim_{\sigma} u_i\,{}^{\sigma}\Delta^{*i}_{mh,h} + Z^0.$$

Hence by 3.4,

$$D^* \sim u_i\,{}^{\sigma}\Delta^{*i}_{mh,h} \text{ in } L.$$

Since D^* was arbitrary in L, it follows that the homology[17] group generated by the cycles ${}^{\sigma}\Delta^{*i}_{mh,h}$ is precisely $H_h(L)$. The number of these Δ^*'s is $R^{\Delta}_{mh} = R_h(K) = R_h(L)$ so that the basis which they constitute is minimal and they are therefore independent relative to homologies in L. Had we regarded D^* as a δ-cycle we would have obtained the same result for the cycles ${}^{\delta}\Delta^{*i}_{mh,h}$.

5.12. *The ρ-cycles of theorem* 5.6 *are independent with respect to ρ-homologies.*

Proof. This is readily verified for the dimension mn. Assume that the theorem is true for the dimensions $mn, \cdots, h+1$. Let us suppose for definiteness that $\rho = \sigma$. Suppose there exists a σ-homology of the form

$$(5.13)\qquad F(\delta Z^0 + Z^*) = x_{ij}\delta T^i\Gamma^j + u_a\,{}^{\sigma}\Delta^a_h + u^s_b\,{}^{\sigma}\Delta^b_{ms,h} + v^t_c\,{}^{\delta}\Delta^c_{mt,h} \sim_{\sigma} 0$$

where $i = 0, \cdots, m-2$, and where $ms - h$ is even, $mt - h$ odd so that all the Δ's are σ-cycles. The cycles ${}^{\sigma}\Delta^a_h (= {}^{\sigma}\Delta^a_{h,h})$ exist only when h is a multiple of m, and since these cycles appear in a term by themselves, we may assume that ms does not assume the value h. Of course $mt \neq h$ because ${}^{\delta}\Delta_{h,h}$ could only represent a δ-cycle. It follows that the cycles in the last two terms are all boundaries of chains ${}^{\sigma}X_{ms,h+1}$, ${}^{\delta}X_{mt,h+1}$ (see (5.5)) hence are ~ 0. Now it is easy to verify that the cycles $\delta\Gamma^j, \cdots, \delta T^{m-2}\Gamma^j$ are independent with respect to ordinary homologies and of course they are also independent of the Δ^a_h's. It follows that the coefficients x_{ij} and u_a in (5.13) must all be 0. Thus (5.13) becomes

$$\begin{aligned} F(\delta Z^0 + Z^*) &= u^s_b\,{}^{\sigma}\Delta^b_{ms,h} + v^t_c\,{}^{\delta}\Delta^c_{mt,h} \\ &= F(u^s_b\,{}^{\sigma}X^b_{ms,h+1} + v^t_c\,{}^{\delta}X^c_{mt,h+1}) \\ &= FY, \text{ say.} \end{aligned}$$

[17] Relative to L.

Since $Y - \delta Z^0 - Z^*$ is a cycle, we may write

$$Y - \delta Z^0 - Z^* \sim \alpha_{ij} T^i \Gamma^j_{k+1} + \beta_k \Delta^k_{k+1}$$

and, on applying σ to both sides, we obtain by 3.5,

$$\sigma Y \sim_\delta \alpha_{ij} T^i \sigma \Gamma^j_{k+1}. \tag{5.14}$$

The index i here ranges from 0 to $m - 1$. But since $T^i\sigma = \sigma$, the right side of (5.14) may be written in the form $\alpha_j \sigma \Gamma^j_{k+1}$; that is, we may assume that the index i assumes only the value 0. Now

$$\sigma Y = u^s_b {}^\sigma\Delta^b_{mh,h+1} + v^t_c \Delta^c_{mt,h+1},$$

all cycles here being δ-cycles. Consequently (5.14) is a δ-homology among the δ-cycles of dimension $h + 1$ which are supposed, by the hypothesis of the induction, to be independent. Hence the u's and v's are 0, and this proves the theorem for the dimension h with $\rho = \sigma$. The modifications necessary for $\rho = \delta$ are obvious.

5.15. $F\Lambda {}^\rho X_{mk,k+1} = \Lambda {}^\rho\Delta^*_{mk,k} \sim 0$ in p.

PROOF. $F {}^\rho X_{mk,k+1} = {}^\rho\Delta_{mk,k} = \lambda {}^\rho X_{mk,k} + {}^\rho\Delta^*_{mk,k}$ (where $\lambda = \rho$ or $\bar{\rho}$) and the desired relations follow from the fact that $\Lambda\lambda = 0$ when $\mathfrak{g} = \mathfrak{m}$.

5.16. *Corollary. Every cycle in l bounds in p.* For since the cycles ${}^\sigma\Delta^{*i}_{mk,k}$ generate the group $H_k(L)$ (Cf. 5.6, II_k), the cycles $\Lambda {}^\sigma\Delta^{*i}_{mk,k}$ generate $H_k(l)$.

5.17. *For $h \geqq t + 2$, $\Lambda {}^\rho X_{mt,h}$ is a cycle in p (to be denoted by ${}^\rho\xi_{mt,h}$).*

PROOF. $F\Lambda X_{mt,h} = \Lambda F X_{mt,h} = \Lambda\Delta_{mt,h-1}$. Since $h - 1 > t$, we have $\Delta^*_{mt,h-1} = 0$. Hence $\Delta_{mt,h-1} = \rho X_{mt,h-1}$ so that $\Lambda\Delta_{mt,h-1} = 0$.

5.18. THEOREM. *Every cycle of dimension h in p is homologous to a linear combination of the cycles ${}^\sigma\xi^b_{ms,h}$, ${}^\delta\xi^c_{mt,h}$, $\Lambda\Gamma^i_h$ where $s + 2 \leqq h$, $ms - h$ odd; $t + 2 \leqq h$, $mt - h$ even.*

PROOF. Let γ_h be a cycle in p. Then $\Lambda'\gamma_h$ is a δ-cycle in P and we may write $\Lambda'\gamma_h = \sigma C_h$, $\Lambda C_h = \gamma_h$. We have by theorem 5.6,

$$\sigma C_h \sim_\delta x_i \sigma \Gamma^i_h + u^s_b {}^\sigma\Delta^b_{ms,h} + v^t_c {}^\delta\Delta^c_{mt,h}, \tag{5.19}$$

where $ms - h$ is odd, $mt - h$ even, all cycles now being δ-cycles. Since m is odd, $mh - h$ is even and therefore $s \neq h$, and hence the Δ^* of every ${}^\sigma\Delta^b_{ms,h}$ which can appear in (5.19) must be 0. Moreover, no ${}^\delta\Delta^{*c}_{mh,h}$ can appear with non-zero coefficient; for otherwise (5.19) would imply that some linear combination of the cycles ${}^\delta\Delta^{*c}_{mh,h}$ is ~ 0 on L, by 3.4, contradicting theorem 5.6, II_h . Hence every $v^h_c = 0$. Since $mh - (h - 1)$ is odd (m being odd) it follows that $t \neq h - 1$. Hence we may assume that t only ranges over values $\leqq h - 2$. Suppose now that some coefficient u^{h-1}_b is $\neq 0$. Then (5.19) can be written

$$0 \sim_\delta \sigma(C_h + x_i \Gamma^i + u^{h-1}_b X^b_{m(h-1),h} + \cdots) = \sigma Y$$

say, where the dots denote a sum of X's whose boundary Δ's do not contain cells of L. These boundary Δ's are therefore carried into 0 by Λ and consequently

$$F\Lambda Y = \Lambda FY = \Lambda u_b^{h-1\,\sigma}\Delta^b_{m(h-1),h-1} = u_b^{h-1}\Lambda^{\,\sigma}\Delta^{*b}_{m(h-1),h-1}. \tag{5.20}$$

Now by 4.2, ΛY is a cycle mod l in p and $\Lambda Y \sim 0 \bmod l$ in p, say $\Lambda Y = Fd \bmod l$, or $Fd = \Lambda Y - c$ ($d \subset p$, $c \subset l$). Then $F\Lambda Y = Fc$ so that $F\Lambda Y \sim 0$ in l. Thus the last term in (5.20) is ~ 0 in l and hence $u_b^{h-1\,\sigma}\Delta^{*b}_{m(h-1),h-1} \sim 0$ in L which by theorem 5.6, II_h is impossible. We conclude therefore that u_b^s can only be different from 0 when $s \leqq h-2$ and therefore (5.19) may be written in the form

$$0 \sim_\delta \sigma(C_h + x_i\Gamma^i_h + u_b^{s\,\sigma}X^b_{ms,h} + v_c^{t\,\delta}X^c_{mt,h}) = \sigma Z,$$

say, where $s+2 \leqq h$, $t+2 \leqq h$. By 4.2, we have $\Lambda Z \sim 0 \bmod l$ in p; that is,

$$\gamma_h + x_i\Lambda\Gamma^i_h + u_b^{s\,\sigma}\xi^b_{ms,h} + v_c^{t\,\delta}\xi^c_{mt,h} \sim 0 \bmod l. \tag{5.21}$$

The expression on the left is however an absolute cycle, and hence it is $\sim$ to a cycle on l. But every cycle on l is ~ 0 in p (Corollary 5.16). Hence the "mod l" may be omitted in (5.21), and the proof is now complete.

5.22. THEOREM. *The cycles in theorem* 5.18 *are independent with respect to homologies in p.*

PROOF. Consider a homology of the form

$$\eta = u_j\Lambda\Gamma^j_h + v_b^{s\,\sigma}\xi^b_{ms,h} + w_c^{t\,\delta}\xi^c_{mt,h} \sim 0.$$

Then from §4, we have $\Lambda'\eta \sim_\delta 0$ in P; i.e.

$$u_j\sigma\Gamma^j_h + v_b^{s\,\sigma}\Delta^b_{ms,h} + w_c^{t\,\delta}\Delta^c_{mt,h} \sim_\delta 0.$$

From theorem I_h (5.6) with $\rho = \delta$, such a relation can only hold if all coefficients vanish.

Theorems 5.6, 5.12, 5.18, and 5.22 describe completely the structure of the groups $H^\sigma_h(P)$, $H^\delta_h(P)$ and $H_h(p)$. They are direct sums of cyclic groups of order m and their ranks can be obtained explicitly in terms of the Betti numbers of K. Consider for example the groups $H_h(p)$. The Betti number $R_h(p)$ equals the number of cycles in the basis described by theorem 5.18. To determine this number, we observe first that since $ms - h$ is odd, s can not assume the value $h-2$ (m being odd). Therefore, since $s \leqq h-2$ and $ms \geqq h$, the values which s may assume are

$$h-3,\ h-5,\ \cdots,\ s_0$$

where s_0 is the smallest integer which is $\geqq h/m$ and of opposite parity from h. Similarly the values which t may assume are

$$h-2,\ h-4,\ \cdots,\ t_0$$

where t_0 is the smallest integer which is $\geqq h/m$ and of same parity as h. Consequently, s and t together may assume the values

$$h - 2, h - 3, \cdots, (h/m)$$

where (h/m) is the smallest integer $\geqq h/m$. Now for each value of s, the number of ${}^{\sigma}\xi_{ms,h}$'s equals the number of ${}^{\sigma}\Delta_{ms,h}$'s, namely $R^{\Delta}_{ms} = R_s(K)$. Similarly the number of ${}^{\sigma}\xi_{mt,h}$'s equals $R^{\Delta}_{mt} = R_t(K)$. Consequently the total number of ξ's is

$$R^{\xi}_h = R_{h-2}(K) + R_{h-3}(K) + \cdots + R_{(m/h)}(K). \tag{5.23}$$

The number of $\Lambda\Gamma^i_h$'s is R^{Γ}_h and it is easy to see that

$$R^{\Gamma}_h = \begin{cases} \dfrac{1}{m}[R_h(P) - R_{h/m}(K)] & \text{when } h \equiv 0 \bmod m \\ \dfrac{1}{m} \cdot R_h(P) & \text{when } h \not\equiv 0 \bmod m \end{cases} \tag{5.24}$$

where $R_h(P) = \sum_{i_1+i_2+\cdots+i_m=h} R_{i_1}(K)R_{i_2}(K) \cdots R_{i_m}(K)$.

We have, then, the result:

5.25. THEOREM. *When* $\mathfrak{g} = \mathfrak{m}$ *and* m *is prime, the* h*-dimensional Betti number of the cyclic product* p_{mn} *is* $R^{\xi}_h + R^{\Gamma}_h$ *where* R^{ξ}_h *is given by* (5.23) *and* R^{Γ}_h *by* (5.24).

It is of interest to remark that the Betti numbers of p modulo l are also expressible in terms of the (absolute) Betti numbers of K. Let ${}^{\rho}\xi^{a}_{ms,h} = \Lambda\, {}^{\rho}X^{a}_{ms,h}$. As we have seen, ${}^{\rho}\xi^{a}_{ms,h}$ is a cycle when $h \geqq s + 2$. We now observe that if $h = s + 1$, ${}^{\rho}\xi^{a}_{ms,h}$ is a cycle mod l. For

$$F\xi_{ms,s+1} = F\Lambda X_{ms,s+1} = \Lambda\, {}^{\rho}\Delta_{ms,s} = \Lambda\, {}^{\rho}\Delta^{*}_{ms,s} = 0 \bmod l.$$

From this it follows that the $\Lambda\Gamma$'s and ξ's which satisfy the description in theorem 5.18 with $s + 2$ and $t + 2$ replaced by $s + 1$ and $t + 1$, constitute a set of cycles mod l. We assert that these cycles are linearly independent with respect to homologies in mod l. For suppose that a linear combination of them with coefficients not all zero is ~ 0 mod l. It follows readily that the same linear combination—call it $\mathfrak{L}$—of the corresponding cycles $\sigma\Gamma^i_h$, ${}^{\sigma}\Delta^{b}_{ms,h}$, ${}^{\delta}\Delta^{c}_{mt,h}$ is $\sim_{\delta} 0$ mod L. Thus there exists a relation of the form

$$F(\sigma Z^0 + Z^*) = \mathfrak{L}(\sigma\Gamma^i, {}^{\sigma}\Delta^{b}_{ms,h}, {}^{\delta}\Delta^{c}_{mt,h}) + Y^*.$$

Now the $\sigma\Gamma$'s and Δ's are each of the form σY^0 and therefore from 3.4 we have $F\sigma Z^0 = \mathfrak{L}$ and hence $\mathfrak{L} \sim_{\delta} 0$; this is only possible if all the coefficients of $\mathfrak{L}$ are zero, which proves our assertion. It can be shown, by an easy modification of the proof of 5.18 that the cycles in question form a basis for $H_h(p; l)$. These cycles are precisely the cycles of theorem 5.18 plus the cycles ${}^{\sigma}\xi_{m(h-1),h}$. (Note that since m is odd, so is $m(h - 1) - h$ as required; there can be no cycles of the form ${}^{\delta}\xi_{m(h-1),h}$ for then $m(h - 1) - h$ would have to be even.) The number

of such cycles is $R^{\Delta}_{m(h-1)}(P) = R_{h-1}(K)$. We have therefore the theorem that *$H_h(p; l)$ is the direct sum of $R_h(p; l)$ cyclic groups of order m where $R_h(p; l) = R_h(p) + R_{h-1}(K)$.*

6. A generalization

Let us consider a complex P_n satisfying the conditions of section 3, which is not a product complex, but which does have, like P_{mn}, a homology basis of the form $T^i\Gamma^j_h$, Δ^a_h with $T\Delta = \Delta$, the elements of each basis being linearly independent. Under certain conditions, the analysis of the situation relative to P_{mn}, T with m prime carries over to the more general P_n, T even though we no longer assume m to be prime. As before we consider the sequences

$$^{\rho}\Delta^c_g,\ ^{\rho}\Delta^c_{g,g-1},\cdots \qquad (\rho = \delta \text{ or } \sigma). \tag{6.1}$$

If $^{\rho}\Delta^{*c}_g$ does not bound in K, let $\mathrm{r}(g, c, \rho) = g$; if $^{\rho}\Delta^{*c}_g$ bounds in K, let $\mathrm{r}(g, c, \rho)$ be the dimension of the first cycle of dimension less than g in (6.1) whose Δ^* fails to bound in L. For a P_{mn}, $\mathrm{r}(g, c, \rho)$ is defined only for values of g which are multiples of m, and if $R^{\Delta}_{ms} \neq 0$, we have (theorem 5.6, II_h)

$$\mathrm{r}(ms, c, \rho) = s \qquad (c = 1, \cdots, R^{\Delta}_{ms}).$$

Consider the set $^{\sigma}D_h$ of σ-cycles

$$\begin{aligned} &^{\sigma}\Delta^a_{k,h} && (k - h \text{ even}), \\ &^{\delta}\Delta^b_{l,h} && (l - h \text{ odd}), \end{aligned} \tag{6.2}$$

where k, a range over all values for which $\mathrm{r}(k, a, \sigma)$ is defined and equal to h, or, as we shall say, (k, a) ranges over all the solutions of the equation $\mathrm{r}(k, a, \sigma) = h$; similarly (b, l) is to range over all solutions of $\mathrm{r}(l, b, \delta) = h$. Let $^{\delta}D_h$ consist of the δ-cycles defined by interchanging the words odd and even in (6.2), and let $^{\rho}D^*_h$ consist of the Δ^*'s of the Δ's in $^{\rho}D_h$. In the proof of theorem 5.18, it was essential to know that the cycles of $^{\sigma}D^*_h$, as well as those of $^{\delta}D^*_h$, were linearly independent with respect to homologies in L. In the more general situation the independence of the cycles $^{\rho}D^*_h$ does not seem to be a necessary consequence of the topology of the situation and therefore will be postulated.

Returning now to theorem 5.18 we see that the cycles $^{\sigma}\xi^a_{ms,h}$ can be described as the cycles $^{\sigma}\xi^a_{k,h}$ when (k, a) ranges over the solutions of the relations

$$2 + \mathrm{r}(k, a, \sigma) \leqq h,\ k \geqq h,\ k - h \equiv 1 \pmod 2 \tag{6.3}$$

and the cycles $^{\delta}\xi^b_{mt,h}$ as the cycles $^{\delta}\xi^b_{l,h}$ when (l, b) ranges over the solutions of

$$2 + \mathrm{r}(l, b, \delta) \leqq h,\ l \geqq h,\ l - h \equiv 0 \pmod 2. \tag{6.4}$$

Thus if we denote the number of solutions of (6.3) and (6.4) by S^{σ}_h, S^{δ}_h, theorem 5.25 takes the form: $R_h(p) = R^{\Gamma}_h + S^{\sigma}_h + S^{\delta}_h$. Except for one modification, this is the form in which the theorem carries over to the more general

case. The modification consists in adding to the expression for $R_h(p)$ the term T_h which denotes the number of solutions (in c) of the equation

$$\mathrm{r}(h, c, \sigma) = h. \tag{6.5}$$

To prove this, we first state the corresponding generalization of theorems 5.6 and 5.12.

6.6. *Let T be a simplicial homeomorphism of P_n into itself of period m. Assume that for a fixed h and for $h \leqq k \leqq n$ each $H_k(P)$, modulo m, is generated by linearly independent cycles of the form $T^i\Gamma_k^j$, Δ_k^a, $(i = 0, \cdots, m-1; j = 1, \cdots, R_k^\Gamma$; $a = 1, \cdots, R_k^\Delta$; $T\Delta = \Delta)$. If the cycles of each set ${}^\sigma D_k^*$ as well as those of ${}^\delta D_k^*$ are linearly independent relative to homologies in L, H_h^σ is the direct sum of the cyclic groups of order m generated by the σ-cycles*

$$\delta T^i \Gamma_h^j, \ {}^\sigma\Delta_{k,h}^a, \ {}^\delta\Delta_{l,h}^b \qquad (i = 0, \cdots, m-2; \Delta_{h,h} = \Delta_h)$$

where (k, a) ranges over the solutions of

$$\mathrm{r}(k, a, \sigma) \geqq h, \qquad k - h \equiv 0 \pmod 2 \tag{6.7}$$

and (l, b) ranges over the solutions of

$$\mathrm{r}(l, b, \delta) \geqq h \qquad l - h \equiv 1 \pmod 2; \tag{6.8}$$

H_h^δ is the direct sum of the cyclic groups generated by

$$\sigma\Gamma_h^j, \ {}^\sigma\Delta_{k,h}^a, \ {}^\delta\Delta_{l,h}^b$$

when (k, a) ranges over the solutions of (6.7) with 0 replaced by 1, and (l, b) ranges over the solutions of (6.8) with 1 replaced by 0.

The proof is exactly like that of theorems 5.6 and 5.12 and may be omitted.

We now turn to the modular space p. We must modify slightly the definition of the cycles ξ of 5.17. Let (k, a) be an arbitrary solution of (6.3). Then ${}^\sigma\Delta_{k,h}^a$ is a δ-cycle, hence ${}^\sigma\Delta_{k,h}^a = \sigma {}^\sigma X_{k,h}^a + {}^\sigma\Delta_{k,h}^{*a}$, $FX = {}^\sigma\Delta_{k,h-1}^a$. Since $h - 1 > \mathrm{r}(k, a, \sigma)$, there exists a chain Y^* such that $FY^* = {}^\sigma\Delta_{k,h-1}^{*a}$. Then

$$F\Lambda(X - Y^*) = \Lambda FX - \Lambda FY^* = \Lambda({}^\sigma\Delta_{k,h-1}^a - {}^\sigma\Delta_{k,h-1}^{*a})$$

and the last term equals zero since the cycle in parenthesis is of the form σZ. Consequently $\Lambda(X - Y^*)$ is a cycle, and we take this cycle for ${}^\sigma\xi_{k,h}^a$. Since ${}^\sigma\Delta_{k,h}^{*a} \sim 0$ in L, hence $\sim_\delta 0$, we have ${}^\sigma\Delta_{k,h}^{*a} \sim_\delta \sigma X$. Therefore

$$\Lambda' {}^\sigma\xi_{k,h}^a = \sigma(X - Y^*) = \sigma X \sim_\delta {}^\sigma\Delta_{k,h}^a. \tag{6.9}$$

In the same manner, we define for each solution (l, b) of (6.4) a cycle ${}^\delta\xi_{l,h}^b$ such that

$$\Lambda' {}^\delta\xi_{l,h}^b \sim_\delta {}^\delta\Delta_{l,h}^b. \tag{6.10}$$

Among the integers $1, \cdots, R_h^\Delta$, let those which are solutions of (6.5) be denoted by c, those which are not by d.

6.11. THEOREM. *Under the hypothesis of 6.6, $H_h(p)$ is the direct sum of the*

cyclic groups of order m generated by the R_h^Γ *cycles* $\Lambda\Gamma_h^i$, *the* S_h^σ *cycles* ${}^\sigma\xi_{k,h}^a$, *the* S_h^δ *cycles* ${}^\delta\xi_{l,h}^b$ *and the* T_h *cycles* $\Lambda\Delta_h^{*c}$.

PROOF. The proof that the cycles in question furnish a basis for $H_h(p)$ follows so closely the proof of theorem 5.18 that it may be omitted. It remains to show that the cycles are linearly independent. Suppose there exists a homology

$$\gamma = u_i\Lambda\Gamma_h^i + v_a^k{}^\sigma\xi_{k,h}^a + w_b^l{}^\delta\xi_{l,h}^b + x_c\Lambda\Delta_h^{*c} \sim 0.$$

Then $\Lambda'\gamma \sim_\delta 0$ in P. But since $\Lambda'\Delta^* = \sigma\Delta^* = 0$, it follows from (6.9), (6.10) and theorem 6.6 that the u's, v's and w's must be zero. Thus our homology reduces to, say,

$$0 \sim x_c\Lambda\Delta_h^{*c} = F(d).$$

Let D be a chain in K such that $\Lambda D = d$, $\sigma D = \Lambda' d$. We have $F(\sigma D) = \Lambda'(x_c\Lambda\Delta_h^{*c}) = 0$ so that σD is a δ-cycle. Hence FD is a σ-cycle and we may then write, say,

$$FD = x_c\Delta_h^{*c} - \delta Z^0 + Z^*.$$

But since $F\Lambda D = \Lambda x_c\Delta_h^{*c}$ and $\Lambda\delta = 0$, it follows that $Z^* = 0$. Thus $x_c\Delta_h^{*c} \sim \delta Z^0$. From theorem 6.6, we have a relation of the form

$$\delta Z^0 - \delta u_{ij}T^i\Gamma_h^j \sim_\sigma y_s^k{}^\sigma\Delta_{k,h}^a + z_t^l{}^\delta\Delta_{l,h}^t. \tag{6.12}$$

If in the expression on the right each Δ be replaced by its Δ^*, the resulting cycle is ~ 0 on L (by 3.4). Those of the Δ^*'s which are not ~ 0 in L are distinct members of ${}^\rho D_h^*$ and are therefore linearly independent in L. Hence their coefficients must vanish. Hence in (6.12) the coefficients of the cycles ${}^\rho\Delta_{h,h}^c$ $(= \Delta_h^c)$ are zero. If we observe that $\Delta_{k,h} \sim 0$ when $k > h$, and that ρ-homology implies ordinary homology, it follows from (6.12) that δZ^0, and therefore $x_c\Delta_h^{*c}$, is homologous to a linear combination of the cycles $T^i\Gamma_h^j$ and the cycles Δ_h^d. By precisely the same argument, $x_c\Delta_h^{0c}(= x_c(\Delta_h^c - \Delta_h^{*c}))$ is homologous to a linear combination of the same sort. Therefore $x_c\Delta_h^c$ also is homologous to a linear combination of the cycles $T^i\Gamma_h^j$ and Δ_h^d. But since the cycles Δ_h^c, Δ_h^d, $T^i\Gamma_h^j$ are linearly independent, the x's must vanish. This concludes the proof.

Let Q_h^σ, Q_h^δ be the number of solutions of (6.3), (6.4), respectively, with the 2 replaced by a 1. Then in a similar manner, we can prove

6.13. THEOREM. *$H_h(p; l)$ is the direct sum of $R_h(p; l)$ cyclic groups of order m where $R_h(p; l) = R_h + Q_h^\sigma + Q_h^\delta$ (under the hypotheses of* 6.6).

Another result whose proof we omit is the following:

6.14. *Under the same hypotheses, $H_h(L)$ is the direct sum of $R_h(L)$ cyclic groups of order m where $R_h(L)$ is the number of solutions of*

$$\mathrm{r}(g, a, \delta) = h, \qquad g - h \equiv 0 \pmod 2$$

plus the number of solutions of

$$\mathrm{r}(j, b, \sigma) = h, \qquad j - h \equiv 1 \pmod 2.$$

Example. Let P be a torus with angular coördinates θ, ϕ and let T be defined by $\theta' = -\theta, \phi' = \phi$. Obviously p is homeomorphic to the plane region bounded by two concentric circles. As a basis for $H_1(P)$ (modulo 2) we may take the curves Δ^1, Δ^2 given by $\theta = 0$ and $\phi = 0$ respectively. There are no Γ's. We find that $\mathrm{r}(2, 1, \rho) = 1$ $\mathrm{r}(1, 1, \rho) = 1$, $\mathrm{r}(1, 2, \rho) = 0$ (note that $\delta = \sigma$ when $m = 2$). Hence for $h = 1, 2$ there are no solutions for the relations $2 + \mathrm{r}(l, a, \rho) \leqq h, l \geqq h$. Thus $S_h^a = S_h^3 = 0$ $(h = 1, 2)$. There are no solutions for $\mathrm{r}(2, a, \rho) = 2$ but there is one solution for $\mathrm{r}(1, a, \rho) = 1$. Thus $T_2 = 0$, $T_1 = 1$ and hence $R_2(p) = 0$, $R_1(p) = 1$.

INSTITUTE FOR ADVANCED STUDY,
BARNARD COLLEGE, COLUMBIA UNIVERSITY.

REFERENCES

(1) G. D. BIRKHOFF AND P. A. SMITH, *Structure analysis of surface transformations*, Journal de Math. 7 (1928) pp. 345–379.
(2) S. LEFSCHETZ, *Topology*, American Math. Soc. Coll. Pub. XII, New York, 1930.
(3) P. A. SMITH, *The topology of involutions*, Proc. Nat. Acad. Sci. 19 (1933) pp. 612–618.
(4) M. RICHARDSON, *On the homology characters of symmetric products*, Duke Math. Journal 1 (1935) pp. 50–69. Correction, ibid. 3 (1937) p. 382.
(5) M. RICHARDSON, *The relative connectivities of symmetric products*, Bull. Amer. Math. Soc. 41 (1935) pp. 528–534.
(6) J. NIELSEN, *Die Struktur Periodische Transformationen von Flächen*, Kgl. Danske Videnskabernes Selskab. Math.-fysiske meddelelsen. XV, 1 (1937).
(7) P. A. SMITH, *Transformations of finite period*, Annals of Math. 39 (1938) pp. 127–164.
(8) M. RICHARDSON, *Special homology groups*, Proc. Nat. Acad. Sci. 24 (1938) pp. 21–23.

ANNALS OF MATHEMATICS
Vol. 39, No. 3, July, 1938

ON THE REGULAR REPRESENTATIONS OF ALGEBRAS*

BY C. NESBITT

(Received May 9, 1938)

Introduction.[1] As this paper is the expression of my work as a student under Prof. R. Brauer, I wish at the outset to acknowledge my deep indebtedness to my former teacher.

For an associative algebra $\mathfrak{a}$, having a unit element and lying over a given field K, two regular representations $\mathfrak{R}$ and $\mathfrak{S}$ are defined. As we shall be concerned with the absolutely irreducible constituents of the regular representations we shall assume that K is algebraically closed. The decomposition of $\mathfrak{R}$ and $\mathfrak{S}$ is our first goal.

Let $\mathfrak{B}$ be a representation of an algebra $\mathfrak{a}$ by linear transformations of a vector space V. If V is the direct sum of two vector spaces V_1 and V_2, both invariant under $\mathfrak{B}$, then $\mathfrak{B}$ is decomposable. By adapting the co-ordinate system in V to the decomposition, we obtain $\mathfrak{B}$ in the form

$$\mathfrak{B} = \begin{pmatrix} \mathfrak{B}_1 & 0 \\ 0 & \mathfrak{B}_2 \end{pmatrix}$$

where $\mathfrak{B}_1$ operates in V_1 and $\mathfrak{B}_2$ in V_2. In the case that the representations $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are themselves decomposable we write them in the same form as $\mathfrak{B}$ above, and continue this process until we finally obtain the splitting of $\mathfrak{B}$ into indecomposable constituents. By breaking up every indecomposable constituent into its irreducible parts we obtain the splitting of $\mathfrak{B}$ into irreducible constituents.[2] The indecomposable constituents are uniquely determined if equivalent representations are considered as equal;[3] that the similar fact holds for the irreducible parts is well known.

We denote the non-equivalent indecomposable constituents of $\mathfrak{R}$ by $\mathfrak{U}_1$, $\mathfrak{U}_2, \cdots, \mathfrak{U}_k$ the non-equivalent indecomposable constituents of $\mathfrak{S}$ by $\mathfrak{V}_1, \mathfrak{V}_2,$

---

* The following comprised a thesis written under the direction of Prof. Brauer and accepted by the University of Toronto, May 1937.

[1] A brief account of this work is given in a joint note by R. Brauer and the writer, Proc. Nat. Acad. Sci. Vol. 23 (1937). A paper by R. Brauer is forthcoming. For a paper by T. Nakayama, *Some studies on regular representations, induced representations and modular representations*, see Ann. Math. 39 (1938), p. 361. For a fourth paper by T. Nakayama and the writer, *Note on Symmetric Algebras*, see Ann. Math. 39 (1938), p. 659. We shall refer to these papers as B. N. Proc., B., N. Ann. and N. N. respectively.

[2] See, for instance, H. Weyl, *The Theory of Groups and Quantum Mechanics*, London (1931) pp. 121–122 for the definitions of irreducibility and equivalence.

[3] W. Krull, Math. Zeit. 23, p. 161 (1925). Cf. also R. Brauer-I. Schur, Sitz. Ber. Berlin, (1930), p. 209.

$\cdots, \mathfrak{B}_l$, and the totality of non-equivalent irreducible representations of $\mathfrak{a}$ by $\mathfrak{F}_1, \mathfrak{F}_2, \cdots, \mathfrak{F}_m$. Combining the powerful method used by Cartan[4] together with new methods, we obtain a number of connections between the $\mathfrak{U}_\kappa$, $\mathfrak{B}_\lambda$ and $\mathfrak{F}_\mu$. In particular, the numbers k, l, and m are equal, and the $\mathfrak{B}_\kappa$, $\mathfrak{U}_\kappa$ may be enumerated in such a way that $\mathfrak{F}_\kappa$ appears as first constituent in $\mathfrak{B}_\kappa$, and as last constituent in $\mathfrak{U}_\kappa$, that is

$$\mathfrak{B}_\kappa = \begin{pmatrix} \mathfrak{F}_\kappa & 0 \\ * & * \end{pmatrix}, \qquad \mathfrak{U}_\kappa = \begin{pmatrix} * & 0 \\ * & \mathfrak{F}_\kappa \end{pmatrix}.$$

These one-to-one correspondences between $\mathfrak{B}_\kappa$ and $\mathfrak{F}_\kappa$, $\mathfrak{U}_\kappa$ and $\mathfrak{F}_\kappa$ are uniquely determined.

We denote by $c_{\kappa\lambda}$ the number of irreducible constituents of $\mathfrak{U}_\kappa$ that are equivalent to $\mathfrak{F}_\lambda$. Then $c_{\lambda\kappa}$ gives the corresponding number for $\mathfrak{B}_\kappa$. We shall call these $c_{\kappa\lambda}$ the Cartan invariants for $\mathfrak{a}$. The Cartan invariants are significant numbers, and several characterizations of them are given.

In §8 attention is confined to algebras for which the two regular representations are equivalent. Frobenius has given a necessary and sufficient condition for this equivalence, and we shall call such algebras Frobenius algebras.[5] In the case of a Frobenius algebra the totality $\mathfrak{U}_1, \mathfrak{U}_2, \cdots, \mathfrak{U}_k$ of indecomposable constituents of $\mathfrak{R}$ coincides with the totality of indecomposable constituents of $\mathfrak{S}$, but $\mathfrak{U}_\kappa$ is not necessarily equivalent to $\mathfrak{B}_\kappa$ and $c_{\kappa\lambda}$ may be different from $c_{\lambda\kappa}$. There exists a sub-class of the Frobenius algebras, namely the symmetric algebras, for which $\mathfrak{U}_\kappa$ and $\mathfrak{B}_\kappa$ are equivalent, and $c_{\kappa\lambda} = c_{\lambda\kappa}$ for all κ, λ. Among well known types of algebras which belong to the symmetric class are the semi-simple algebras, and the group rings of finite groups even in the case where the latter are not semi-simple, i.e. if the characteristic of the underlying field is a prime which divides the order of the group.

1. Let $\mathfrak{a}$ be an associative algebra, possessing a unit element ϵ and lying over an algebraically closed field K, and let $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ be a basis for $\mathfrak{a}$. For ever α in $\mathfrak{a}$ we have equations

$$\epsilon_i \cdot \alpha = \sum_{j=1}^{n} r_{ij}^{(\alpha)} \epsilon_j, \tag{1}$$

$$\alpha \cdot \epsilon_i = \sum_{j=1}^{n} s_{ji}^{(\alpha)} \epsilon_j, \tag{2}$$

where the coefficients $r_{ij}^{(\alpha)}$ and $s_{ij}^{(\alpha)}$ lie in K. The two regular representations $\mathfrak{R}$ and $\mathfrak{S}$ of $\mathfrak{a}$ are obtained by associating the matrices $R(\alpha) = (r_{ij}^{(\alpha)})$, $S(\alpha) = (s_{ij}^{(\alpha)})$

[4] E. Cartan, *Les groupes bilinéaires et les systèmes de nombres complexes*, Annales de Toulouse, 12 B, p. 1 (1898). We shall refer to this paper as C.

[5] Cf. G. Frobenius, *Theorie der hyperkomplexen Grössen* I, Sitz. Ber. Berlin, (1903). The writer, in collaborating with T. Nakayama, adopted the term Frobeniusean algebra, but now, quailing before our critics, we return to simply Frobenius algebra.

with α. The matrices of $\mathfrak{R}$ are of degree n, and $\mathfrak{R}$ is a faithful[6] representation of $\mathfrak{a}$. The similar facts hold for $\mathfrak{S}$.$\mathfrak{R}$ is called the first, and $\mathfrak{S}$, the second, regular representation.

In what follows for the matrix $B(\epsilon_i) = (b_{jk}^{(\epsilon_i)})$ which in any representation $\mathfrak{B}$ of $\mathfrak{a}$ is associated with the basis element ϵ_i we use a contracted notation, namely $B(\epsilon_i) = B_i = (b_{jk}^i)$. Then in the formation of $S(\epsilon_i) = S_i$ we would consider equations of form

$$\epsilon_i \epsilon_j = \sum_{k=1}^{n} s_{kj}^i \epsilon_k \tag{3}$$

but thinking in terms of the first regular representation (3) becomes

$$\epsilon_i \epsilon_j = \sum_{k=1}^{n} r_{ik}^j \epsilon_k, \tag{4}$$

Comparing the right hand members of (3) and (4) we obtain the relation

$$r_{ik}^j = s_{kj}^i. \tag{5}$$

We require some additional definitions. Let $\mathfrak{B}$ and $\mathfrak{C}$ be any two representations of $\mathfrak{a}$ which associate the matrices $B(\alpha)$ and $C(\alpha)$ with the element α of $\mathfrak{a}$. $\mathfrak{B}$ is *intertwined* with $\mathfrak{C}$ is there exists a matrix $P \neq 0$ for which

$$B(\alpha)P = PC(\alpha) \text{ for all } \alpha \text{ in } \mathfrak{a}$$

holds. We say that $\mathfrak{B}$ is *h times intertwined* with $\mathfrak{C}$ if among all such matrices P there are exactly h that are linearly independent, and we indicate this relation by setting $h = I(\mathfrak{B}, \mathfrak{C})$.

THEOREM 1. *If $\mathfrak{R}$ and $\mathfrak{S}$ are the first and the second regular representations of an associative algebra $\mathfrak{a}$, and $\mathfrak{B}$ is a representation of $\mathfrak{a}$, of degree b, which does not contain the 0-representation as constituent, then*

$$b = I(\mathfrak{R}, \mathfrak{B}) = I(\mathfrak{B}, \mathfrak{S}).^7$$

We shall prove the first statement, $b = I(\mathfrak{R}, \mathfrak{B})$.

Since $\mathfrak{B}$ is a representation of $\mathfrak{a}$, we have from (1)

$$B_i \cdot B(\alpha) = \sum_{j=1}^{n} r_{ij}^{(\alpha)} B_j,$$

or (6)

$$(b_{kl}^i)(b_{kl}^{(\alpha)}) = \sum_{j=1}^{n} r_{ij}^{(\alpha)} (b_{kl}^j).$$

---

[6] The term faithful is used to denote a representation which is isomorphic to $\mathfrak{a}$; cf. H. Weyl, *Generalized Riemann Matrices and Factor Sets*, Ann. Math. 37 (1936) p. 711.

[7] This theorem is the work of R. Brauer (cf. B.). I give it here completely as both the results and the method of proof are basic in the following. This theorem and its applications are typical of the viewpoint of R. Brauer in these questions while T. Nakayama uses the language of homomorphisms. In this connection see N. N., Lemma 1 and Remark 1.

Taking the (r, s) element on each side of (6), we obtain

$$\sum_{u=1}^{b} b_{ru}^{i} b_{us}^{(\alpha)} = \sum_{j=1}^{n} r_{ij}^{(\alpha)} b_{rs}^{j}. \tag{7}$$

We form the matrix $P_r = (b_{rj}^{i})_{ij}$ by taking together the r^{th} row from each of the n matrices $B_1, B_2, \cdots, B_n$. Then P_r has n rows and b columns. Giving r the values $1, 2, \cdots, b$ we obtain b such-matrices. Further each P_r intertwines $\mathfrak{R}$ and $\mathfrak{B}$ for, on examination, we see that the (i, s) element of $R(\alpha)P_r$ is $\sum_{j=1}^{n} r_{ij}^{(\alpha)} b_{rs}^{j}$ and that the (i, s) element of $P_r B(\alpha)$ is $\sum_{u=1}^{b} b_{ru}^{i} b_{us}^{(\alpha)}$, but these are equal by (7).

Hence $I(\mathfrak{R}, \mathfrak{B}) \geqq b$ provided that the b matrices $P_1, P_2, \cdots, P_b$ are linearly independent. But this in fact is the case, for suppose $\sum_{m=1}^{b} c_m P_m$ with not all $c_m = 0$ vanishes. Then

$$\sum_{m} c_m b_{mj}^{i} = 0 \tag{8}$$

holds for all $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, b$. Forming the one-rowed matrix $C = (c_1, c_2, \cdots, c_b)$ we may write (8) as

$$0 \cdot C = C \cdot B_i.$$

Thus the 0-representation and $\mathfrak{B}$ are intertwined by $C \neq 0$, and by Schur's Lemma[8] $\mathfrak{B}$ would then contain the 0-representation as constituent contrary to our hypothesis.

We may show on the other hand that $b \geqq I(\mathfrak{R}, \mathfrak{B})$. For let $V = (v_{ij})$ be any matrix intertwining $\mathfrak{R}$ and $\mathfrak{B}$. Then

$$R_t V = V B_t \qquad (t = 1, 2, \cdots, n)$$

that is

$$\sum_{j=1}^{n} r_{ij}^{t} v_{jk} = \sum_{u=1}^{b} v_{iu} b_{uk}^{t}. \tag{9}$$

Let V_j be the matrix $(v_{j1}, v_{j2}, \cdots, v_{jb})$ having one row and b columns. Then (9) may be written as

$$\sum_{j=1}^{n} r_{ij}^{t} V_j = V_i B_t. \tag{10}$$

Let us choose the basis element ϵ_1 to be the unit element of $\mathfrak{a}$, then setting $i = 1$ we have $r_{1j}^{t} = \delta_{tj} = \begin{cases} 0, & t \neq j \\ 1, & t = j \end{cases}$, and from (10)

$$V_1 B_t = V_t. \tag{11}$$

[8] Cf. I. Schur, Sitz. Ber. Berlin (1905), p. 406.

If we know the row V_1 in V, we are then able to complete the whole matrix V. Starting with a number of choices for V_1, say

$$V_1^{(1)}, V_1^{(2)}, \cdots, V_1^{(l)}, \tag{12}$$

we construct the corresponding

$$V^{(1)}, V^{(2)}, \cdots, V^{(l)}. \tag{13}$$

We see immediately by (11) that a linear dependence of the matrices (12) implies the same dependence for the matrices (13). Then the number of linearly independent matrices intertwining $\mathfrak{R}$ and $\mathfrak{B}$ is at most equal to the number of linearly independent matrices with one row and b columns. It follows that $I(\mathfrak{R}, \mathfrak{B}) \leqq b$, and comparing with the former inequality we obtain $I(\mathfrak{R}, \mathfrak{B}) = b$.

In the proof of the second statement, $b = I(\mathfrak{B}, \mathfrak{S})$, we would define matrices Q_r formed by taking the r^{th} column from each of the matrices B_j, thus obtaining matrices $(b_{ir}^j)_{ij}$ with b rows and n columns. These b matrices Q_r form a complete linearly independent set of matrices intertwining $\mathfrak{B}$ and $\mathfrak{S}$, the term complete implying that any matrix P which intertwines $\mathfrak{B}$ with $\mathfrak{S}$ may be expressed as a linear combination of the $Q_r (r = 1, 2, \cdots, b)$. The proof is similar in every detail to that above.

As an immediate application of the method of Theorem I we have

COROLLARY 1. *The matrices* $T_m = (r_{im}^j)$, $(m = 1, 2, \cdots, n)$ *form a complete linearly independent set of matrices intertwining* $\mathfrak{R}$ *with* $\mathfrak{S}$.

Again, from the method of the theorem, and the relation (5) we obtain the well known result[9]

COROLLARY 2. *The matrices* S_j', $(j = 1, 2, \cdots, n)$ *form a complete linearly independent set of matrices which intertwine* $\mathfrak{R}$ *with itself, that is, which commute with every element of* $\mathfrak{R}$. *The matrices* R_j' *have the same property with respect to* $\mathfrak{S}$.

2. In analyzing the determinants of the regular representations, Cartan selected a particular basis system the use of which exhibited the factor structure of these determinants.[10] We shall employ the same basis system together with more recent concepts, in particular, that of intertwining matrices, to study the representation structure of the regular representations.

We now replace the elements of the abstract algebra $\mathfrak{a}$ by the matrices of a faithful representation $\mathfrak{A}$ of $\mathfrak{a}$, and shall in the following consider $\mathfrak{A}$ to be our algebra. Let $\mathfrak{F}_1, \mathfrak{F}_2, \cdots, \mathfrak{F}_k$ be the distinct irreducible representations of $\mathfrak{a}$. We assume $\mathfrak{A}$ to be split into its irreducible constituents, thus

$$\mathfrak{A} = \begin{pmatrix} \mathfrak{F}_{\rho 1} & & & 0 \\ \mathfrak{A}_{21} & \mathfrak{F}_{\rho 2} & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \mathfrak{A}_{n1} & & \cdots & \mathfrak{F}_{\rho n} \end{pmatrix} \tag{14}$$

[9] Cf. H. Weyl, (op. cit. in footnote 6) p. 715.

[10] Cf. C.

where each $\mathfrak{F}_{\rho_i}$ is equivalent to some $\mathfrak{F}_j$, and each $\mathfrak{F}_j$ appears at least once since $\mathfrak{A}$ is a faithful representation of $\mathfrak{a}$. $\mathfrak{A}$ may be expressed as the direct sum of its radical $\mathfrak{N}$ and a semisimple algebra $\mathfrak{L}$, such that by proper choice of coordinates[11]

$$\mathfrak{L} = \begin{pmatrix} \mathfrak{F}_{\rho 1} & & & \\ & \mathfrak{F}_{\rho 2} & & 0 \\ & & \ddots & \\ 0 & & & \mathfrak{F}_{\rho n} \end{pmatrix}.$$

$\mathfrak{L}$ may in turn be written as the direct sum of simple algebras $\mathfrak{L}_1, \mathfrak{L}_2, \cdots, \mathfrak{L}_k$ where $\mathfrak{L}_\lambda$ contains only constituents $\mathfrak{F}_\lambda$, thus

$$\mathfrak{L}_\lambda = \begin{pmatrix} 0 & & & & \\ & \ddots & & 0 & \\ & & \mathfrak{F}_\lambda & & \\ & & & \ddots & \\ & 0 & & & \mathfrak{F}_\lambda \\ & & & & & \ddots \\ & & & & & & 0 \end{pmatrix}. \tag{15}$$

In particular, the unit element E_λ of $\mathfrak{L}_\lambda$ has a unit matrix of proper degree in each of the places where $\mathfrak{F}_\lambda$ appears in $\mathfrak{A}$, and has zero elsewhere. At once follow the relations

$$E_\kappa E_\lambda = 0, \qquad \kappa \neq \lambda, \qquad E_\kappa E_\kappa = E_\kappa, \tag{16}$$

$$E_1 + E_2 + \cdots + E_k = E, \tag{17}$$

where E is the unit element of $\mathfrak{A}$. Any matrix A of $\mathfrak{A}$ is such that

$$A = EAE = \sum_{\kappa,\lambda} E_\kappa A E_\lambda = \sum_{\kappa,\lambda} A_{\kappa\lambda}. \tag{18}$$

The matrix $A_{\kappa\lambda} = E_\kappa A E_\lambda$ belongs to the algebra $\mathfrak{A}$, since it is a product of matrices contained in $\mathfrak{A}$.

An element M of $\mathfrak{A}$ is said to be of type (μ, ν) if

$$E_\mu M E_\nu = M. \tag{19}$$

Then $A_{\kappa\lambda}$ is of type (κ, λ) for from (16)

$$E_\kappa A_{\kappa\lambda} E_\lambda = E_\kappa E_\kappa A E_\lambda E_\lambda = E_\kappa A E_\lambda = A_{\kappa\lambda}.$$

In particular, E_κ is of type (κ, κ). The product of an element M of type (κ, λ) and N of type (μ, ν) is

$$\begin{aligned} M \cdot N &= E_\kappa M E_\lambda E_\mu N E_\nu = 0, \ \lambda \neq \mu \\ &= E_\kappa M E_\lambda \cdot N E_\nu = E_\kappa M \cdot N E_\nu, \ \lambda = \mu \quad \text{(cf. (19), (16))} \end{aligned} \tag{20}$$

and is, therefore, of type (κ, ν).

[11] See, for instance, L. E. Dickson, *Algebras and their Arithmetics*, Chicago, 1923, or M. Deuring, *Algebren*, Ergebn. Math. 4. (1935).

We denote by $E_\mu(mn)$ the matrix which has 1 for the (mn) coefficient of the irreducible constituent $\mathfrak{F}_\mu$ of $\mathfrak{A}$, and zero elsewhere. That such a matrix belongs to $\mathfrak{L}$, and hence to $\mathfrak{A}$, follows from the Burnside Theorem.[12] We may easily verify the relations

$$E_\kappa = \sum_i E_\kappa(ii),$$

$$E_\mu(mn) \cdot E_\nu(pq) = \begin{cases} 0 & \text{if either } \mu \neq \nu \text{ or } n \neq p, \\ E_\mu(mq) & \text{if both } \mu = \nu \text{ and } n = p. \end{cases} \tag{21}$$

Then to select a basis for elements of type (κ, λ) we form the product

$$P = E_\kappa(ij) A_{\kappa\lambda} E_\lambda(pq). \tag{22}$$

Now, by its construction, $A_{\kappa\lambda}$ can have non-zero coefficients only in blocks at the intersection of those rows which in $\mathfrak{A}$ (cf. (14)) contain $\mathfrak{F}_\kappa$, and of those columns which in $\mathfrak{A}$ contain $\mathfrak{F}_\lambda$. We denote these blocks by $T_{\kappa\lambda}$, and the degree of $\mathfrak{F}_\mu$ by f_μ. Then each of these blocks $T_{\kappa\lambda}$ has f_κ rows and f_λ columns. In the formation of the product (22) each such block is multiplied on the left by a matrix of degree f_κ having 1 for its (ij) coefficient, and zero elsewhere, and on the right by a matrix of degree f_λ having 1 for its (pq) element and its remaining coefficients zero. The effect is to make the (jp) element of the original block the (iq) element in the product block, and to give zero elsewhere.

We choose (iq) to be (11) so that the resulting blocks can have non-zero coefficients only in their upper left corners. From these matrices

$$E_\kappa(1j) A_{\kappa\lambda} E_\lambda(p1) \tag{23}$$

we shall select a basis system after the following consideration.

We denote by $\mathfrak{N}^r$ the set containing all elements of form $\sum N_{\mu_1} N_{\mu_2} \cdots N_{\mu_r}$ where the N_{μ_e} are elements contained in the radical $\mathfrak{N}$. For a sufficiently high value t, $\mathfrak{N}^t = 0$. It is readily verified that these $\mathfrak{N}^r$ form invariant subalgebras of $\mathfrak{A}$. If we denote $\mathfrak{A}$ by $\mathfrak{N}^0$, we have

$$\mathfrak{A} = \mathfrak{N}^0 \supset \mathfrak{N}^1 \supset \mathfrak{N}^2 \supset \cdots \supset \mathfrak{N}^{l-1} \supset \mathfrak{N}^l = 0.$$

We say that an element A *belongs* to $\mathfrak{N}^h$ if h is the largest value such that $\mathfrak{N}^h$ contains A. In particular, the semisimple subalgebra $\mathfrak{L}$ belongs to $\mathfrak{N}^0$. We denote matrices which belong to $\mathfrak{N}^h$ by $A^{(h)}$, $B^{(h)} \cdots (h = 0, 1, 2, \cdots, l)$. Then $A^{(s)} \cdot B^{(r)} = C^{(t)}$ where $t \geqq r + s$.

In choosing a basis for the matrices (23) we first select as many basis elements as possible from the set $\mathfrak{N}^{l-1}$, then as many as possible from $\mathfrak{N}^{l-2}$ besides those already taken in $\mathfrak{N}^{l-1}$ and so on (at each stage we require, of course, linear independence). We thus obtain a basis

$$B^1_{\kappa\lambda}, B^2_{\kappa\lambda}, \cdots, B^m_{\kappa\lambda} \tag{24}$$

[12] See, for instance, B. L. van der Waerden, *Moderne Algebra*, II, p. 183.

for the matrices (23) such that any matrix of (23) which belongs to $\mathfrak{N}^\rho$ is expressible as a linear combination of the matrices (24) which belong to the sets $\mathfrak{N}^\mu$, $\mu \geqq \rho$. Using Cartan's notation we shall denote by $c_{\kappa\lambda}$ the number m of matrices in the set (24). To each type (μ, ν) there is thus associated a number $c_{\mu\nu}$; in the following these $c_{\mu\nu}$ are of considerable significance.[13]

From the matrices (24) we form the associated matrices

$$(25) \qquad \begin{aligned} &\mu = 1, 2, \cdots, c_{\kappa\lambda}, \\ &E_\kappa(i1)B^\mu_{\kappa\lambda}E_\lambda(1q) \quad i = 1, 2, \cdots, f_\kappa, \\ &q = 1, 2, \cdots, f_\lambda, \end{aligned}$$

and denote this set (25) by the symbol $\mathbf{B}_{\kappa\lambda}$. Observe that, by definition, $E_\kappa(11)B^\mu_{\kappa\lambda}E_\lambda(11) = B^\mu_{\kappa\lambda}$. The elements of $\mathbf{B}_{\kappa\lambda}$ are linearly independent, for if

$$\sum_{i,q,\mu} a_{iq\mu}E_\kappa(i1)B^\mu_{\kappa\lambda}E_\lambda(1q) = 0$$

then, multiplying by $E_\kappa(1r)$ on the left, and by $E_\lambda(s1)$ on the right, we have (cf. (21))

$$\sum_\mu a_{rs\mu}B^\mu_{\kappa\lambda} = 0$$

but the $B^\mu_{\kappa\lambda}$ by choice are linearly independent, hence the coefficients $a_{rs\mu}$ are zero. Moreover, let $A_{\kappa\lambda}$ be any matrix of type (κ, λ). Then

$$\begin{aligned} A_{\kappa\lambda} = E_\kappa A_{\kappa\lambda}E_\lambda &= \sum_{j,k} E_\kappa(jj)A_{\kappa\lambda}E_\lambda(kk) \\ &= \sum_{j,k} E_\kappa(j1)\cdot E_\kappa(1j)A_{\kappa\lambda}E_\lambda(k1)\cdot E_\lambda(1k). \end{aligned}$$

But $P = E_\kappa(1j)A_{\kappa\lambda}E_\lambda(k1)$ belongs to the set (23), hence P may be expressed as a linear combination of the elements (24). It follows that $A_{\kappa\lambda}$ is expressible as a linear combination of the elements (25). Accordingly, the set $\mathbf{B}_{\kappa\lambda}$ forms a basis for all matrices of $\mathfrak{A}$ which are of type (κ, λ).

We repeat this method of selection to obtain bases $\mathbf{B}_{\mu\nu}$ for the matrices of each non-trivial type (μ, ν). Taking together the elements of all these sets $\mathbf{B}_{\mu\nu}$ we obtain finally a basis of our algebra $\mathfrak{A}$,

$$(26) \qquad \begin{aligned} &\mu = 1, 2, \cdots, c_{\kappa\lambda}, \\ &\mathbf{B}_{\kappa\lambda}: E_\kappa(i1)B^\mu_{\kappa\lambda}E_\lambda(1j), \quad i = 1, 2, \cdots, f_\kappa, \qquad (\kappa, \lambda = 1, 2, \cdots, k), \\ &j = 1, 2, \cdots, f_\lambda, \end{aligned}$$

for:

1) Every element of $\mathfrak{A}$ is by (18) a linear sum of elements $A_{\kappa\lambda}$, and $A_{\kappa\lambda}$ may be written as a linear sum of elements of the basis $\mathbf{B}_{\kappa\lambda}$ for type (κ, λ).

2) The elements of this set (26) are linearly independent. A linear relation

[13] Cf. C.

which includes elements of types (κ, λ) may be reduced by multiplying on the left by E_κ and on the right by E_λ to a linear relation among the elements of the set $\mathbf{B}_{\kappa\lambda}$ which we have seen above to be linearly independent.

3. We shall now order the elements of the basis (26) in such a way (as we shall see later) that the second regular representation formed with regard to the ordered basis is split into its indecomposable and its irreducible parts.

a) The elements (25) give a basis $\mathbf{B}_{\kappa\lambda}$ for all elements of type (κ, λ). We obtain a basis for all elements of types (μ, λ), λ fixed, $\mu = 1, 2, \cdots, k$ by taking together the elements of all $\mathbf{B}_{\mu\lambda}$ $(\mu = 1, 2, \cdots, k)$. We symbolize the set so formed by B_λ. We thus have k sets $\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_k$.

b) We arrange the elements of $\mathbf{B}_\lambda$ into subsets $\mathbf{B}_\lambda^j$, $j = 1, 2, \cdots f_\lambda$, where $\mathbf{B}_\lambda^j$ contains the elements

$$\begin{aligned} &\kappa = 1, 2, \cdots, k, \\ E_\kappa(i1)B_{\kappa\lambda}^\mu E_\lambda(1j),\ &i = 1, 2, \cdots, f_\kappa, \\ &\mu = 1, 2, \cdots, c_{\kappa\lambda}. \end{aligned}$$

c) Finally we arrange the elements of $\mathbf{B}_\lambda^j$ according to the sets $\mathfrak{N}^r$ (see §2) to which they belong. We first select the matrices $B_{\kappa\lambda}^\mu$ of the set (24) which belong to $\mathfrak{N}^0$. There is only one, and from (15) we may take it to be the matrix $E_\lambda(11)$ having, as its only non-zero coefficients, 1 in the upper left corner of each $\mathfrak{F}_\lambda$ appearing in $\mathfrak{A}$. We take the elements associated with $E_\lambda(11)$

$$E_\lambda(i1)E_\lambda(11)E_\lambda(1j) = E_\lambda(ij) \qquad (i = 1, 2, \cdots, f_\lambda) \tag{27}$$

as the first group of elements in the set $\mathfrak{B}_\lambda^j$ and arrange these according to ascending values of i. Next we select the $B_{\kappa\lambda}^\mu$ which belong to $\mathfrak{N}^1$, and arrange their associated elements as the second group in $\mathbf{B}_\lambda^j$, and so on.

3′. The analogous ordering of the basis elements (26) for the purpose of obtaining a reduced form for the first regular representation is as follows.

a′) We form the set $\bar{\mathbf{B}}_\kappa$ by taking together elements of all $\mathbf{B}_{\kappa\mu}$ $(\mu = 1, 2, \cdots, k)$.

b′) We arrange the elements of $\bar{\mathbf{B}}_\kappa$ into subsets $\bar{\mathbf{B}}_\kappa^i$, $i = 1, 2, \cdots, f_\kappa$, where $\bar{\mathbf{B}}_\kappa^i$ contains the elements

$$\begin{aligned} &\lambda = 1, 2, \cdots, k, \\ E_\kappa(i1)B_{\kappa\lambda}^\mu E_\lambda(1j),\ &j = 1, 2, \cdots, f_\lambda, \\ &\mu = 1, 2, \cdots, c_{\kappa\lambda}. \end{aligned}$$

c′) In the set $\bar{\mathbf{B}}_\lambda^i$ we take first the elements which belong to the highest power of $\mathfrak{N}$, then those belonging to the next highest power of $\mathfrak{N}$, and so on.

4. We shall now form the second regular representation of $\mathfrak{A}$ with respect to the basis (26) ordered in the manner described in §3. In the following when we

speak of the product of an element A or $\mathfrak{A}$ with a basis element, we shall mean that A is multiplied on the right by the basis element. Then the product of A and an element from the set $\mathbf{B}_\lambda$ (cf. (§3, a)) is by (18) and (20) expressible as a linear combination of the elements of $\mathbf{B}_\lambda$. As a result, the second regular representation (cf. (2)) splits in the form

$$\mathfrak{S} = \begin{pmatrix} \mathfrak{P}_1 & & & \\ & \mathfrak{P}_2 & & 0 \\ & & \ddots & \\ & 0 & & \mathfrak{P}_k \end{pmatrix} \tag{28}$$

where the block $\mathfrak{P}_\lambda$ is associated with the set $\mathbf{B}_\lambda$.

Let us next consider more precisely the product of A and an element $P = E_\kappa(i1)B^\mu_{\kappa\lambda}E_\lambda(1j)$ from the set $\mathbf{B}^j_\lambda$ (cf. §3, b)). We multiply P with A in steps. We have first from (18) and the definitions of $E_\kappa(i1)$, $B^\mu_{\kappa\lambda}$ that

$$A \cdot E_\kappa(i1)B^\mu_{\kappa\lambda} = \sum_{\tau,\rho,l} a\begin{pmatrix} \mu, & \kappa, & \lambda, & i \\ \tau, & \rho, & l & \end{pmatrix} E_\tau(l1)B^\rho_{\tau\lambda}, \tag{29}$$

where the $a\begin{pmatrix} \mu, & \kappa, & \lambda, & i \\ \tau, & \rho, & l & \end{pmatrix}$ are the coefficients in the expression of the product as a linear combination of the basis elements (26). Now multiplying on the right by $E_\lambda(1j)$ throughout (29) we have

$$A \cdot E_\kappa(i1)B^\mu_{\kappa\lambda}E_\lambda(1j) = \sum_{\tau,\rho,l} a\begin{pmatrix} \mu, & \kappa, & \lambda, & i \\ \tau, & \rho, & l & \end{pmatrix} E_\tau(l1)B^\rho_{\tau\lambda}E_\lambda(1j) \tag{30}$$

where the coefficients $a\begin{pmatrix} \mu, & \kappa, & \lambda, & i \\ \tau, & \rho, & l & \end{pmatrix}$ are the same as in (29), and are thus independent of j. Then corresponding to the grouping of the elements of the set $\mathbf{B}_\lambda$ into the subsets $\mathbf{B}^j_\lambda$ ($j = 1, 2, \cdots, f_\lambda$) we obtain a splitting of the block $\mathfrak{P}_\lambda$ (cf. (28)) into f_λ parts, each of which has the same coefficients by (30). Naming these new parts $\mathfrak{W}_\lambda$, we have $\mathfrak{S}$ now in the form

$$\mathfrak{S} = \begin{pmatrix} E_{f_1} \times \mathfrak{W}_1 & & & \\ & E_{f_2} \times \mathfrak{W}_2 & & 0 \\ & & \ddots & \\ & 0 & & E_{f_k} \times \mathfrak{W}_k \end{pmatrix} \tag{31}$$

where E_f is a unit matrix of degree f, and $E_f \times \mathfrak{W}$ is the Kronecker product of the matrices E_f and $\mathfrak{W}$.

We shall now consider a fixed $\mathfrak{W}$, and shall simplify our notation slightly. We leave away the factor $E_\lambda(1j)$ in P, and instead of the general relation (30) we consider the formulas which express $A \cdot P = A \cdot E_\kappa(i1)B^\mu_{\kappa\lambda}$ by the elements $E_\tau(l1)B^\rho_{\tau\lambda}$ (cf. (29)).

We shall now take account of the arrangement described in §3, c). In the

relation (29) let us suppose $P = E_\kappa(i1)B^\mu_{\kappa\lambda}$ belongs to $\mathfrak{N}^r$. Now, in general, A will be expressible as a linear sum of basis elements belonging to various sets $\mathfrak{N}^t$ from $\mathfrak{N}^0$ to $\mathfrak{N}^{l-1}$. The product of those components of A which belong to $\mathfrak{N}^s$, $s > 0$ and P will (see §2) belong to $\mathfrak{N}^t$, $t > r$, so may be written as a linear sum of basis elements belonging to sets $\mathfrak{N}^\mu$, $\mu \geqq t > r$, and thus give rise to coefficients below the main diagonal in $\mathfrak{W}_\lambda$. On the other hand, the product of those components of A which belong to $\mathfrak{N}^0$ and P may contain terms belonging to $\mathfrak{N}^r$, and consequently give coefficients in a block $\mathfrak{C}_r$ in the main diagonal of $\mathfrak{W}_\lambda$. It follows that

$$\mathfrak{W}_\lambda = \begin{pmatrix} \mathfrak{C}_0 & & \\ & \mathfrak{C}_1 & 0 \\ & \ddots & \ddots \\ & * & \mathfrak{C}_\sigma \end{pmatrix} \tag{32}$$

where the block $\mathfrak{C}_\rho$ appears at the intersection of those rows and columns of $\mathfrak{W}_\lambda$ which are associated with basis elements belonging to $\mathfrak{N}^\rho$.

Let us examine these blocks $\mathfrak{C}_\rho$ more precisely. We shall first consider the product of A and a basis element $P = E_\lambda(i1)E_\lambda(11) = E_\lambda(i1)$ from the set (27). Expressing A as a sum of the basis elements (26) we have (cf. (27))

$$A = \text{terms in } \mathfrak{N}^t + \sum_{\mu,m,n} f^{(\mu)}_{mn} E_\mu(mn) \tag{33}$$

where $t > 0$ and the $f^{(\mu)}_{mn}$ are coefficients from the matrix $F_\mu(A)$ which corresponds to A in the irreducible representations $\mathfrak{F}_\mu$ of $\mathfrak{A}$. Then, from the relations (21)

$$A \cdot E_\lambda(i1) = \text{terms in } \mathfrak{N}^t + \sum_m f^{(\lambda)}_{mi} E_\lambda(m1). \tag{34}$$

Giving i the values $1, 2, \cdots, f_\lambda$, and comparing with (32) we see that the first constituent $\mathfrak{C}_0$ of $\mathfrak{W}_\lambda$ is the irreducible representation $\mathfrak{F}_\lambda$ of $\mathfrak{A}$, that is,

$$\mathfrak{W}_\lambda = \begin{pmatrix} \mathfrak{F}_\lambda & 0 \\ * & * \end{pmatrix}. \tag{35}$$

We have already seen that $\mathfrak{W}_\lambda$ appears f_λ times as constituent of $\mathfrak{S}$ where f_λ is the degree of $\mathfrak{F}_\lambda$ (cf. (31)).

From these facts it follows that $\mathfrak{W}_\lambda$ is indecomposable. For suppose $\mathfrak{W}_\lambda$ can be decomposed into two parts $\mathfrak{W}'_\lambda$ and $\mathfrak{W}''_\lambda$. By a suitable transformation of co-ordinates we may arrange that $\mathfrak{W}'_\lambda$ and $\mathfrak{W}''_\lambda$ are split into irreducible constituents. Let their first irreducible constituents be $\mathfrak{F}'_\lambda$ and $\mathfrak{F}''_\lambda$ respectively. As irreducible part of the representation $\mathfrak{S}$, $\mathfrak{F}'_\lambda$ must be equivalent to one of the irreducible representations $\mathfrak{F}_\mu (\mu = 1, 2, \cdots, k)$ of $\mathfrak{A}$. If $\mathfrak{F}'_\lambda$ is equivalent to $\mathfrak{F}_\kappa$, $\kappa \neq \lambda$, then $\mathfrak{F}_\kappa$ appears as first constituent of both $\mathfrak{W}'_\lambda$ and $\mathfrak{W}_\kappa$ so that (cf. (31)) on permuting rows and columns we may write $\mathfrak{S}$ as

$$\mathfrak{S} \sim \begin{pmatrix} E_{f_\lambda + f_\kappa} \times \mathfrak{F}_\kappa & 0 \\ * & * \end{pmatrix}.$$

Then the matrices

$$(x_1, x_2, \cdots, x_{f_\lambda+f_\kappa}, 0) \times E_{f_\kappa}$$

where the x_i are arbitrary elements from the field K, intertwine $\mathfrak{F}_\kappa$ and $\mathfrak{S}$. Consequently, $I(\mathfrak{F}_\kappa, \mathfrak{S}) > f_\kappa$ which is in contradiction to Theorem 1. The only possibility left is that both $\mathfrak{F}'_\lambda$ and $\mathfrak{F}''_\lambda$ are equivalent to $\mathfrak{F}_\lambda$. But since in this case $\mathfrak{F}_\lambda$ appears as first constituent of both $\mathfrak{W}'_\lambda$ and $\mathfrak{W}''_\lambda$ from the same argument as above $I(\mathfrak{F}_\lambda, \mathfrak{S}) \geqq 2f_\lambda$ which again is a contradiction.

Denoting the non-equivalent indecomposable constituents of $\mathfrak{S}$ by $\mathfrak{B}_1$, $\mathfrak{B}_2, \cdots, \mathfrak{B}_l$ we have

THEOREM 2. *The number l of distinct indecomposable constituents $\mathfrak{B}_\lambda$ or $\mathfrak{S}$ is equal to the number k of distinct irreducible representations $\mathfrak{F}_\mu$ of $\mathfrak{A}$. We may arrange the $\mathfrak{B}_\lambda$ and the $\mathfrak{F}_\mu$ in such a manner that $\mathfrak{B}_\lambda$ contains $\mathfrak{F}_\lambda$ as first irreducible constituent $(\lambda = 1, 2, \cdots, k)$, i.e.*

$$\mathfrak{B}_\lambda = \begin{pmatrix} \mathfrak{F}_\lambda & 0 \\ * & * \end{pmatrix}. \tag{36}$$

It is not possible to find another splitting of $\mathfrak{B}_\lambda$ of the form (36) *in which there stands another completely reducible representation $\mathfrak{H}$ at the place of $\mathfrak{F}_\lambda$.*

The first part follows by enumerating the $\mathfrak{B}_\lambda$ so that $\mathfrak{B}_\lambda$ is $\mathfrak{W}_\lambda$. The last statement results from the fact that all possibilities for $\mathfrak{H}$ other than $\mathfrak{F}_\lambda$ itself, lead by the above form of argument to a contradiction to Theorem 1.

We shall say that a representation $\mathfrak{B}$ is split into its top series of Loewy maximal completely reducible parts if

$$\mathfrak{B} = \begin{pmatrix} \mathfrak{B}_1 & & & \\ * & \mathfrak{B}_2 & 0 & \\ \cdot & \cdot & \cdot & \\ * & \cdot & \cdot & \mathfrak{B}_s \end{pmatrix}$$

where the $\mathfrak{B}_\nu$ are completely reducible, and $\mathfrak{B}_1$ cannot be replaced by a completely reducible part of higher degree, further, if $\mathfrak{B}_1$ is fixed, then $\mathfrak{B}_2$ cannot be replaced by a completely reducible part of higher degree, and so on. Analogously we shall say that $\mathfrak{B}$ is split into its bottom series of Loewy maximal completely reducible parts if

$$\mathfrak{B} = \begin{pmatrix} \overline{\mathfrak{B}}_t & & & \\ \cdot & \cdot & 0 & \\ \cdot & \cdot & \overline{\mathfrak{B}}_2 & \\ * & \cdot & * & \overline{\mathfrak{B}}_1 \end{pmatrix}$$

where $\overline{\mathfrak{B}}_1$ is completely reducible and maximal, $\overline{\mathfrak{B}}_2$ is completely reducible, and maximal if $\overline{\mathfrak{B}}_1$ is held fixed, and so forth. If equivalent representations are considered as equal, then the $\mathfrak{B}_\nu$ and $\overline{\mathfrak{B}}_\mu$ are uniquely determined.[14] We now show

[14] A. Loewy, *Über die vollständig reduziblen Gruppen, die zu einer Gruppe linearer homogener Substitutionen gehören*, Trans. Amer. Math. Soc. 6 (1905) p. 504.

THEOREM 3. *By arranging the basis elements* $E_\kappa(i1)B^\mu_{\kappa\lambda}(\lambda$ *fixed) with first those which belong to* $\mathfrak{N}^0$, *next those which belong to* $\mathfrak{N}^1$, *and so forth, the indecomposable constituent* $\mathfrak{B}_\lambda$ *of* $\mathfrak{S}$ *splits into its top series of Loewy maximal completely reducible parts.*

We have already seen that under this arrangement of the basis elements $E_\kappa(i1)B^\mu_{\kappa\lambda}$, $\mathfrak{B}_\lambda$ has the form (32) where the block $\mathfrak{C}_r$ appears at the intersection of those rows and columns of $\mathfrak{B}_\lambda$ which are associated with basis elements belonging to $\mathfrak{N}^r$.

i) The part $\mathfrak{C}_r$ is completely reducible. Let

$$B^\rho_{\kappa\lambda}, B^{\rho'}_{\kappa'\lambda}, \cdots, B^\nu_{\tau\lambda} \tag{37}$$

and their associated basis elements be those matrices of the set $\mathbf{B}^j_\lambda$ (see §3, b)) which belong to $\mathfrak{N}^r$. Consider the product of A and $P = E_\kappa(i1)B^\rho_{\kappa\lambda}$. Again taking A in the form (33), we obtain from the relations (21)

$$A \cdot E_\kappa(i1)B^\rho_{\kappa\lambda} = \text{terms in } \mathfrak{N}^s + \sum_m f^{(\kappa)}_{mi} E_\kappa(m1)B^\rho_{\kappa\lambda} \tag{38}$$

where $s > r$ and the $f^{(\kappa)}_{mi}$ are coefficients from $F_\kappa(A)$. Observe that no basis element belonging to $\mathfrak{N}^r$ other than those associated with $B^\rho_{\kappa\lambda}$ appears in the right member of (38). Giving i the values $1, 2, \cdots, f_\lambda$ we see that $\mathfrak{F}_\kappa$ appears as a constituent of $\mathfrak{C}_r$. By application of the argument for each element of the set (37) it follows that

$$\mathfrak{C}_r = \begin{pmatrix} \mathfrak{F}_\kappa & & \\ & \mathfrak{F}_{\kappa'} & 0 \\ & & \ddots \\ & 0 & \mathfrak{F}_\tau \end{pmatrix}$$

ii) We shall now show that the $\mathfrak{C}_r$ give the top series of Loewy maximal completely reducible parts of $\mathfrak{B}_\lambda$ (cf. (32)). We have already seen (cf. Theorem 2) that $\mathfrak{C}_0 = \mathfrak{F}_\lambda$ is the uniquely determined first maximal completely reducible part of $\mathfrak{B}_\lambda$. Suppose $\mathfrak{C}_0, \mathfrak{C}_1, \cdots, \mathfrak{C}_{r-1}$ are maximal (in the Loewy sense), then we shall show that $\mathfrak{C}_r$ is also maximal.

It can be shown[15] that if, when $\mathfrak{C}_0, \mathfrak{C}_1, \cdots, \mathfrak{C}_{r-1}$ are fixed, a larger completely reducible part can be obtained in place of $\mathfrak{C}_r$, it may be done by means of similarity transformations of two types on $\mathfrak{B}_\lambda$. Each row and column of $\mathfrak{B}_\lambda$ is associated with a definite basis element (cf. (2)). The transformations are (a) Multiplying the row p (associated with a basis element belonging to $\mathfrak{N}^r$) by a constant c and adding to row q (associated with a basis element belonging to $\mathfrak{N}^s$, $s > r$), and corresponding operations for columns, viz. multiplying column q by $-c$ and adding to column p; (b) Rearranging rows and columns which are associated with basis elements belonging to $\mathfrak{N}^s$, $s > r$.

[15] This is a theorem of R. Brauer and will appear in B.

For the second regular representation[16] this means that the basis elements which belong to $\mathfrak{N}^r$ may be changed by the addition of elements which belong to a higher power of $\mathfrak{N}$, and that perhaps the basis elements belonging to the sets $\mathfrak{N}^s$, $s > r$ are taken in another arrangement. If then we would have

$$\mathfrak{B}_\lambda = \begin{pmatrix} \mathfrak{C}_0 & & & & \\ & \ddots & & & \\ & & \mathfrak{C}_{r-1} & & 0 \\ & & & \begin{pmatrix} \mathfrak{C}_r' & 0 \\ 0 & * \end{pmatrix} & \\ & * & & & \ddots \end{pmatrix}$$

it would mean that on multiplying an element N in $\mathfrak{N}$ by a basis element belonging to $\mathfrak{N}^r$ we would obtain a linear combination of the basis elements following those belonging to $\mathfrak{N}^r$ but the first basis elements immediately after those of $\mathfrak{N}^r$ would not occur (cf. (2) and §2). If M is such a basis element it belongs to $\mathfrak{N}^{r+a}$, $a > 0$ hence $M = \sum_\nu X_\nu Y_\nu$, X_ν in $\mathfrak{N}^a$, Y_ν in $\mathfrak{N}^r \cdot Y_\nu$ may be written as (α) a linear combination of the basis elements belonging to $\mathfrak{N}^r$ plus (β) a linear combination of basis elements which belong to sets $\mathfrak{N}^s$, $s > r$. By the above remark the product of X_ν and (α), when expressed in terms of the basis elements, cannot contain the basis element M, neither can the product of X_ν and the part (β), as this latter product is contained in $\mathfrak{N}^{a+s}$ and $s > r$. Then $\sum_\nu X_\nu Y_\nu$, expressed as a sum of basis elements, does not contain M, which gives a contradiction. This completes Theorem 3.

We say that an irreducible constituent $\mathfrak{F}$ of a representation $\mathfrak{B}$ has the multiplicity h if h of the irreducible constituents of $\mathfrak{B}$ are equivalent to $\mathfrak{F}$. We may now easily obtain

THEOREM 4. *$\mathfrak{F}_\kappa$ has multiplicity $c_{\kappa\lambda}$ as irreducible constituent in $\mathfrak{B}_\lambda$.*

PROOF. If in the relation (38) we use in turn each of the $c_{\kappa\lambda}$ independent $B^\mu_{\kappa\lambda}$ (κ, λ fixed) we obtain for each such element an irreducible constituent $\mathfrak{F}_\kappa$ in $\mathfrak{B}_\lambda$. Thus $\mathfrak{F}_\kappa$ appears at least $c_{\kappa\lambda}$ times as constituent in $\mathfrak{B}_\lambda$. But from (33) and the relations (21) we see that the $B^\mu_{\kappa\lambda}(\mu = 1, 2, \cdots, c_{\kappa\lambda})$ and their associated elements $(E_\kappa(i1)B^\mu_{\kappa\lambda})$ are the only basis elements of types $(\nu, \lambda)(\nu = 1, 2, \cdots k)$ which can give rise to the constituent $\mathfrak{F}_\kappa$, and so the statement of the theorem follows.

4′. By forming the first regular representation with respect to the basis elements (26) ordered as in §3′, and applying arguments similar in every detail to those in the preceding section we obtain the analogues of Theorems 2, 3 and 4.

[16] It is to be observed that, if $(\epsilon_1, \epsilon_2, \cdots, \epsilon_n)$ denotes the basis system used in the formation of $\mathfrak{S}$, then transforming $\mathfrak{S}$ into $T^{-1}\mathfrak{S}T$ is equivalent to replacing $(\epsilon_1, \epsilon_2, \cdots, \epsilon_n)$ by $(\epsilon_1, \epsilon_2, \cdots, \epsilon_n)T$.

Thus, denoting the non-equivalent indecomposable constituents of $\mathfrak{R}$ by $\mathfrak{U}_1$, $\mathfrak{U}_2, \cdots, \mathfrak{U}_k$ we have

THEOREM 2′. *The number of distinct indecomposable parts $\mathfrak{U}_\kappa$ of the first regular representation $\mathfrak{R}$ is equal to the number of irreducible representations $\mathfrak{F}_\mu$ of $\mathfrak{A}$. We can enumerate the $\mathfrak{U}_\lambda$ in such a way that $\mathfrak{F}_\lambda$ is the last constituent in $\mathfrak{U}_\lambda$*

$$\mathfrak{U}_\lambda = \begin{pmatrix} * & 0 \\ * & \mathfrak{F}_\lambda \end{pmatrix}. \tag{39}$$

There exists no splitting of $\mathfrak{U}_\lambda$ of the form (39) *in which there stands another completely reducible part in place of $\mathfrak{F}_\kappa$.*

THEOREM 3′. *By arranging the basis elements $B^\mu_{\kappa\lambda}E_\lambda(1j)$ with first those which belong to the highest power of $\mathfrak{N}$, then those belonging to the next highest power of $\mathfrak{N}$ and so on, the indecomposable part $\mathfrak{U}_\kappa$ of $\mathfrak{R}$ splits into its bottom series of Loewy maximal completely reducible parts.*

THEOREM 4′. *$\mathfrak{F}_\lambda$ has multiplicity $c_{\kappa\lambda}$ as irreducible constituent in $\mathfrak{U}_\kappa$.*

5. We have defined the multiplicity of an irreducible constituent of a representation $\mathfrak{B}$ and we may similarly define the multiplicity of an indecomposable constituent of $\mathfrak{B}$. We denote the degree of $\mathfrak{U}_\lambda$ by u_λ, and that of $\mathfrak{B}_\lambda$ by v_λ. From (31) it follows that $\mathfrak{B}_\lambda$ has multiplicity f_λ as indecomposable constituent in $\mathfrak{S}$, and similarly $\mathfrak{U}_\lambda$ has the same multiplicity f_λ in $\mathfrak{R}$. On the other hand, from Theorem 4 and (31) we have that the multiplicity of $\mathfrak{F}_\lambda$ as irreducible constituent in $\mathfrak{S}$ is $\sum_{\mu=1}^{k} c_{\lambda\mu}f_\mu$. But on considering the degrees of $\mathfrak{U}_\lambda$ and of its irreducible constituents we have from Theorem 4′ that $\sum_{\tau=1}^{k} c_{\lambda\tau}f_\tau = u_\lambda$, so that the degree u_λ of the part $\mathfrak{U}_\lambda$ of $\mathfrak{R}$ gives the multiplicity of $\mathfrak{F}_\lambda$ in $\mathfrak{S}$. Similarly from Theorems 4 and 4′ the degree $v_\lambda = \sum_{\tau=1}^{k} c_{\tau\lambda}f_\tau$ of $\mathfrak{B}_\lambda$ gives the multiplicity of $\mathfrak{F}_\lambda$ as irreducible constituent in $\mathfrak{R}$. Moreover,

$$n = \sum_{\rho=1}^{k} u_\rho f_\rho = \sum_{\rho=1}^{k} v_\rho f_\rho = \sum_{\kappa,\lambda=1}^{k} c_{\kappa\lambda}f_\kappa f_\lambda.$$

6. We can now prove

THEOREM 5. *If $\mathfrak{B}$ is a representation of $\mathfrak{A}$, and if the multiplicity of $\mathfrak{F}_\kappa$ as irreducible constituent in $\mathfrak{B}$ is h_λ then*

$$h_\lambda = I(\mathfrak{B}, \mathfrak{B}_\lambda) = I(\mathfrak{U}_\lambda, \mathfrak{B}).^{17}$$

PROOF. We shall show the first statement, $h_\lambda = I(\mathfrak{B}, \mathfrak{B}_\lambda)$. The proof is simple in idea but complicated in notation. We shall use the basis system ordered as in §3, and shall simply apply the method of Theorem 1 to construct h_λ

[17] This theorem was first established by R. Brauer and will be discussed in B. I give here a somewhat less elegant proof than his, but one I wish to make use of in other work.

linearly independent matrices which intertwine $\mathfrak{B}$ with $\mathfrak{B}_\lambda$. Let us, for simplicity, suppose $\lambda = 1$, then by (31)

$$\mathfrak{S} = \begin{pmatrix} \mathfrak{B}_1 & & & \\ & \mathfrak{B}_1 & & 0 \\ & & \cdot & \\ & & & \cdot \\ & 0 & & \mathfrak{B}_1 \\ & & & * \end{pmatrix} \tag{40}$$

where $\mathfrak{B}_1$ appears f_1 times. There is a set of v_1 basis elements from the system (26) associated with the first indecomposable part $\mathfrak{B}_1$, a second set of v_1 basis elements associated with the second part $\mathfrak{B}_1$, and so on. The first f_1 basis elements of these sets are, respectively, by §3, b) and c)

$$E_1(i1); E_1(i2); \cdots ; E_1(if_1) \qquad (i = 1, 2, \cdots, f_1).$$

Now suppose $\mathfrak{B}$ has a constituent $\mathfrak{F}_1$ which contains the intersection of the k^{th} row and k^{th} column of $\mathfrak{B}$.

$$\mathfrak{B} = \begin{pmatrix} \cdot & & \\ & \cdot & 0 \\ & \mathfrak{F}_1 & \\ * & & \cdot \\ & & \cdot \end{pmatrix}. \tag{41}$$

We denote the degree of $\mathfrak{B}$ by b. Let $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ for the moment represent the elements of the basis (26) for $\mathfrak{A}$, and $B(\epsilon_h) = (b_{ij}^h)_{ij}$ the matrix in $\mathfrak{B}$ corresponding to ϵ_h. Then by Theorem 1, the matrix $Q_k = (b_{ik}^j)_{ij}$ of b rows and n columns, formed by taking the k^{th} column from each of the matrices $B(\epsilon_j)$, intertwines $\mathfrak{B}$ with $\mathfrak{S}$. We write Q_k in the form

$$Q_k = (Q_k^{(1)}, Q_k^{(2)}, \cdots, Q_k^{(f_1)}, *) \tag{42}$$

where $Q_k^{(1)}$ contains the first set of v_1 columns of Q_k, $Q_k^{(2)}$ the second set of v_1 columns of Q_k, and so on. Since Q_k intertwines $\mathfrak{B}$ with $\mathfrak{S}$, then from (40) and (42) each $Q_k^{(i)}$ intertwines $\mathfrak{B}$ with $\mathfrak{B}_1$. Now the k^{th} column of $\mathfrak{B}$ is the p^{th} column of the constituent $\mathfrak{F}_1$ in question, $p \leqq f_1$. Let us examine then, in particular, the matrix $Q_k^{(p)}$ which contains the p^{th} set of v_1 columns of Q_k. Consider the basis elements associated with this p^{th} set of v_1 columns. The first f_1 columns of this set are associated, as remarked above, with the basis elements $E_1(ip)$ $(i = 1, 2, \cdots, f_1)$. By definition $E_1(ip)$ has 1 for the (ip) component of the constituent $\mathfrak{F}_1$ of $\mathfrak{A}$ and zero elsewhere. Then the matrix $B(E_1(ip))$, which in $\mathfrak{B}$ corresponds to $E_1(ip)$, has 1 for the (ip) component of the constituent $\mathfrak{F}_1$ of $\mathfrak{B}$ and zero elsewhere in $\mathfrak{F}_1$. As for the last $v_1 - f_1$ basis elements of the p^{th} set, they belong to the radical and hence are represented by zero in $\mathfrak{F}_1$. From this it follows

$$Q_k^{(p)} = \begin{pmatrix} 0 & 0 \\ E_{f_1} & 0 \\ * & * \end{pmatrix} \tag{43}$$

where E_{f_1} is a unit matrix of degree f_1 appearing between the same rows of $Q_k^{(p)}$ as does $\mathfrak{F}_1$ in $\mathfrak{B}$ (cf. (41)).

Thus for each constituent $\mathfrak{F}_1$ that $\mathfrak{B}$ contains we can construct a matrix of the form (43) with b rows and v_1 columns which intertwines $\mathfrak{B}$ with $\mathfrak{B}_1$. Since in each case the unit matrix appearing in the first f_1 columns is between the same rows as is the particular constituent $\mathfrak{F}_1$ in $\mathfrak{B}$, then the matrices so constructed are linearly independent. Hence $\mathfrak{B}$ is at least h_1 times intertwined with $\mathfrak{B}_1$. Similarly

$$I(\mathfrak{B}, \mathfrak{B}_\lambda) \geqq h_\lambda . \tag{44}$$

Then considering the number of times $\mathfrak{B}$ is intertwined with $\mathfrak{S}$ we have since $\mathfrak{B}_\lambda$ has multiplicity f_λ in $\mathfrak{S}$, and $\mathfrak{F}_\lambda$, multiplicity h_λ in $\mathfrak{B}$,

$$I(\mathfrak{B}, \mathfrak{S}) \geqq \sum_{\lambda=1}^{k} h_\lambda f_\lambda = b. \tag{45}$$

By Theorem 1 we may have only the equality sign in (45) and hence also in (44). This completes the proof of the first statement in the theorem, the second follows by a similar argument.

From Theorems 4, 4′ and 5 we obtain the following characterization of the numbers

$$c_{\kappa\lambda} = I(\mathfrak{U}_\kappa , \mathfrak{B}_\lambda) = I(\mathfrak{U}_\lambda , \mathfrak{U}_\kappa) = I(\mathfrak{B}_\lambda , \mathfrak{B}_\kappa).$$

7. Besides the above method of analyzing the regular representations there are others which yield related results.[18] For instance, R. Brauer has shown connections between the indecomposable constituents of a matrix ring B and the irreducible constituents of a matrix ring A of the same degree whose matrices are commutative with every matrix of B.[19] This method is applicable to the regular representations since by Theorem 1, Cor. 2, the transpose of each matrix of $\mathfrak{S}$ is commutative with every matrix of $\mathfrak{R}$. If we take the basis (26) to be arranged as in §3 for the second regular representation we obtain

$$\mathfrak{R}_0 = \begin{pmatrix} \mathfrak{F}_1 \times E_{v_1} & & & 0 \\ & \mathfrak{F}_2 \times E_{v_2} & & \\ & & \ddots & \\ 0 & & & \mathfrak{F}_k \times E_{v_k} \end{pmatrix},$$

$$\mathfrak{S}' = \begin{pmatrix} E_{f_1} \times \mathfrak{B}_1' & & & 0 \\ & E_{f_2} \times \mathfrak{B}_2' & & \\ & & \ddots & \\ 0 & & & E_{f_k} \times \mathfrak{B}_k' \end{pmatrix} \tag{46}$$

[18] In particular, see the more general case discussed in N. Ann.

[19] Cf. R. Brauer, *Über die Darstellung von Gruppen in Galoisschen Feldern*, Actualités scientifiques et industrielles, No. 195 (1935).

where $\mathfrak{R}_0$ is the semisimple sub-system contained in $\mathfrak{R}$. The relations (46) are equivalent to those obtained by applying Brauer's method to the regular representations.

A result of Frobenius is also of interest here.[20] Let $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ be the n basis elements (26) of $\mathfrak{A}$, and put

$$x = x_1\epsilon_1 + x_2\epsilon_2 + \cdots + x_n\epsilon_n, \qquad y = y_1\epsilon_1 + y_2\epsilon_2 + \cdots + y_n\epsilon_n.$$

We denote by $C(x)$, $C(y)$ the matrices representing x and y in any representation $\mathfrak{C}$ of the algebra $\mathfrak{A}$. According to Frobenius' result the determinant of the matrix $(S(x) + R'(y))$ may be written as

$$|S(x) + R'(y)| = \prod_{\kappa,\lambda} \psi_{\kappa\lambda}^{d_{\kappa\lambda}} \tag{47}$$

with

$$\psi_{\kappa\lambda} = \prod_{\alpha,\beta} (u_\kappa^{(\alpha)} + v_\lambda^{(\beta)}) \tag{48}$$

where $u_\kappa^{(\alpha)}$ runs over all characteristic roots of $F_\kappa(x)$ and $v_\lambda^{(\beta)}$ over all roots of $F_\lambda(y)$. Here the $\psi_{\kappa\lambda}$ are distinct irreducible polynomials in $x_1, \cdots, x_n, y_1, \cdots, y_n$. We shall show that the exponents $d_{\kappa\lambda}$ are the Cartan invariants $c_{\kappa\lambda}$ of $\mathfrak{A}$.

We again take the basis elements (26) to be arranged as in §3. Then $R_0(y)$ and $S'(x)$ have the forms given in (46), so that

$$|S(x) + R'(y)| = \prod_{\lambda} |E_{f_\lambda} \times V_\lambda(x) + F'_\lambda(y) \times E_{v_\lambda}|. \tag{49}$$

Choose x and y to be elements of $\mathfrak{A}$ such that

$$F_\lambda(x) = a_\lambda F_{f_\lambda}, \qquad F_\lambda(y) = b_\lambda E_{f_\lambda} \tag{50}$$

where a_λ and b_λ are arbitrary numbers from the field K. The characteristic roots of $|E_{f_\lambda} \times V_\lambda(x) + F'_\lambda(y) \times E_{v_\lambda}|$ are then given by

$$|E_{f_\lambda} \times V_\lambda(x) + b_\lambda E_{f_\lambda v_\lambda} - mE_{f_\lambda v_\lambda}| = 0.$$

It follows that

$$m = b_\lambda + p,$$

where p is a characteristic root of $E_{f_\lambda} \times V_\lambda(x)$. The characteristic roots of $V_\lambda(x)$ are from Theorem 4 and (50), a_κ with multiplicity $c_{\kappa\lambda}f_\kappa$, $\kappa = 1, 2, \cdots, k$. Then the roots of $E_{f_\lambda} \times V_\lambda(x)$ are a_κ with multiplicity $c_{\kappa\lambda}f_\kappa f_\lambda$, $\kappa = 1, 2, \cdots, k$. It follows that

$$|E_{f_\lambda} \times V_\lambda(x) + F'_\lambda(y) \times E_{v_\lambda}| = \prod_{\kappa,\lambda} (a_\kappa + b_\lambda)^{c_{\kappa\lambda}f_\kappa f_\lambda}. \tag{51}$$

But, on the other hand, from (48) and (50)

$$\psi_{\kappa\lambda} = (a_\kappa + b_\lambda)^{f_\kappa f_\lambda}, \tag{52}$$

so that, on comparing (47), (49), (51) and (52)

$$\prod_{\kappa,\lambda} (a_\kappa + b_\lambda)^{c_{\kappa\lambda}f_\kappa f_\lambda} = \prod_{\kappa,\lambda} (a_\kappa + b_\lambda)^{d_{\kappa\lambda}f_\kappa f_\lambda}.$$

[20] G. Frobenius (op. cit. in footnote 8) p. 529.

Since the a_κ , b_λ are arbitrary, we obtain

$$c_{\kappa\lambda} = d_{\kappa\lambda} \qquad (\kappa, \lambda = 1, 2, \cdots, k).$$

8. Let us return for the moment to the abstract algebras, and consider now algebras $\mathfrak{a}$ for which the two regular representations are equivalent assuming as before that $\mathfrak{a}$ possesses a unit element. A necessary and sufficient condition for this equivalence has been given by Frobenius, namely that the parastrophic determinant should not identically vanish.[21] In our notation this condition is

$$(53) \qquad |T| = \left|\sum_{m=1}^{n} T_m z_m\right| = \left|\sum_{m=1}^{n} r_{im}^{j} z_m\right| \not\equiv 0 \text{ in } z_1, z_2, \cdots, z_m,$$

where according to Theorem 1, Cor. 1, T is the most general matrix such that

$$\Re T = T\mathfrak{S}.$$

We shall call such an $\mathfrak{a}$ a Frobenius algebra.

If we consider $\mathfrak{a}$ as an n-dimensional vector space, with basis ϵ_1 , ϵ_2 , $\cdots$, ϵ_n and assume $\Re$ is equivalent to $\mathfrak{S}$, then for some c_1 , c_2 , $\cdots$, c_n

$$(54) \qquad \left|\sum_m T_m c_m\right| = \left|\sum_m r_{im}^{j} c_m\right| \neq 0.$$

It follows that the hyperplane L, consisting of all elements $x = \sum_{m=1}^{n} x_m \epsilon_m$ of $\mathfrak{a}$ with $\sum_{m=1}^{n} c_m x_m = 0$ does not contain a right ideal (except the null ideal). For if L contains a right ideal $\mathfrak{c}$ it contains the principal right ideal generated by an element of $\mathfrak{c}$. Suppose L contains the principal right ideal (ξ) where $\xi = \mu_1\epsilon_1 + \mu_2\epsilon_2 + \cdots + \mu_n\epsilon_n$. From (4)

$$\xi \cdot \epsilon_j = \sum_{i,m=1}^{n} \mu_i r_{im}^{j} \epsilon_m,$$

and for L to contain ξ we must have

$$(55) \qquad \sum_m c_m \sum_i \mu_i r_{im}^{j} = 0,$$

or

$$(56) \qquad \sum_i \mu_i \sum_m r_{im}^{j} c_m = 0,$$

which gives a contradiction to (54) unless all μ_i are zero. Conversely if L: $\sum_m c_m x_m = 0$ does not contain a right ideal, the relation (54) holds and $\Re$ is equivalent to $\mathfrak{S}$. For if (54) were not true there would exist elements μ_i from the underlying field, not all zero, satisfying (56) and hence also (55), so that L would contain the right ideal (ξ) where $\xi = \sum_i \mu_i \epsilon_i$. The condition characterizing

[21] G. Frobenius, op. cit.

Frobenius algebras may thus be given in the form: Not every hyperplane in $\mathfrak{a}$ contains a right ideal (other than the null ideal).[22]

In the case of a Frobenius algebra $\mathfrak{R}$ and $\mathfrak{S}$, due to their equivalence, have the same irreducible constituents, and hence from section 5 $u_\lambda = v_\lambda$ $(\lambda = 1, 2, \cdots, k)$. The totality $\mathfrak{U}_1, \mathfrak{U}_2, \cdots, \mathfrak{U}_k$ of indecomposable constituents of $\mathfrak{R}$ coincides with the totality of indecomposable constituents of $\mathfrak{S}$. But, as the following example shows, $\mathfrak{U}_k$ may not be equivalent to $\mathfrak{B}_\kappa$, and $c_{\kappa\lambda}$ may be different from $c_{\lambda\kappa}$.

We take

$$\mathfrak{A} = \begin{pmatrix} \mathfrak{F}_3 & & & & \\ \mathfrak{C}_1 & \mathfrak{F}_1 & & 0 & \\ & & \mathfrak{F}_1 & & \\ & 0 & \mathfrak{C}_2 & \mathfrak{F}_2 & \\ & & & & \mathfrak{F}_2 \\ & & & \mathfrak{C}_3 & \mathfrak{F}_3 \end{pmatrix}$$

containing all matrices of the above form in which for each of the parts $\mathfrak{F}_1$, $\mathfrak{F}_2$, $\mathfrak{F}_3$, $\mathfrak{C}_1$, $\mathfrak{C}_2$, $\mathfrak{C}_3$ are taken arbitrary elements from the underlying field. Here the irreducible constituents $\mathfrak{F}_\lambda (\lambda = 1, 2, 3)$ are of first degree. The basis elements ordered properly for the first regular representation are then (cf. §§2, 3′) $B_{13}, B_{11}, B_{21}, B_{22}, B_{32}, B_{33}$ where

$$B_{13} = \begin{pmatrix} 0 & & & & & \\ 1 & 0 & & & & \\ & & 0 & & & \\ & & 0 & 0 & & \\ & & & & 0 & \\ & & & & 0 & 0 \end{pmatrix}, \quad B_{11} = \begin{pmatrix} 0 & & & & & \\ 0 & 1 & & & & \\ & & 1 & & & \\ & & 0 & 0 & & \\ & & & & 0 & \\ & & & & 0 & 0 \end{pmatrix},$$

$$B_{21} = \begin{pmatrix} 0 & & & & & \\ 0 & 0 & & & & \\ & & 0 & & & \\ & & 1 & 0 & & \\ & & & & 0 & \\ & & & & 0 & 0 \end{pmatrix}, \quad B_{22} = \begin{pmatrix} 0 & & & & & \\ 0 & 0 & & & & \\ & & 0 & & & \\ & & 0 & 1 & & \\ & & & & 1 & \\ & & & & 0 & 0 \end{pmatrix},$$

$$B_{32} = \begin{pmatrix} 0 & & & & & \\ 0 & 0 & & & & \\ & & 0 & & & \\ & & 0 & 0 & & \\ & & & & 0 & \\ & & & & 1 & 0 \end{pmatrix}, \quad B_{33} = \begin{pmatrix} 1 & & & & & \\ 0 & 0 & & & & \\ & & 0 & & & \\ & & 0 & 0 & & \\ & & & & 0 & \\ & & & & 0 & 1 \end{pmatrix}.$$

[22] We mean by a hyperplane in $\mathfrak{a}$ an $(n-1)$-dimensional vector subspace of $\mathfrak{a}$.

On forming the first regular representation we obtain $\mathfrak{R} = \mathfrak{A}$. Taking the basis elements in the order B_{33}, B_{13}, B_{11}, B_{21}, B_{22}, B_{32} for the second regular representation (cf. §3) we have $\mathfrak{S} = \mathfrak{A}$ also. Evidently, $\mathfrak{R}$ and $\mathfrak{S}$ are equivalent. However, $\mathfrak{U}_1 = \begin{pmatrix} \mathfrak{F}_3 & 0 \\ \mathfrak{L}_1 & \mathfrak{F}_1 \end{pmatrix}$ is not equivalent to $\mathfrak{V}_1 = \begin{pmatrix} \mathfrak{F}_1 & 0 \\ \mathfrak{L}_2 & \mathfrak{F}_2 \end{pmatrix}$, and the $c_{\kappa\lambda}$ are not symmetric, e.g. $c_{13} = 1$, $c_{31} = 0$.

We shall call an algebra symmetric if the following condition is satisfied: There exists a hyperplane in $\mathfrak{a}$ which contains all commutator elements $\alpha\beta - \beta\alpha$ (α, β elements of $\mathfrak{a}$) but does not contain a right ideal (except the null ideal). It follows at once that every symmetric algebra is a Frobenius algebra.

Suppose $L: \sum_{m=1}^{n} c_m x_m = 0$ is the given hyperplane. Since L does not contain a right ideal then (cf. the above) the matrix $T = (\sum_m r^j_{im} c_m)$ has determinant different from zero and transforms $\mathfrak{R}$ into $\mathfrak{S}$.

Let $\alpha = \sum_{i=1}^{n} x_i \epsilon_i$, $\beta = \sum_{i=1}^{n} y_i \epsilon_i$ where the x_i, y_i are arbitrary elements from the underlying field. We find

$$\begin{aligned} \alpha\beta - \beta\alpha &= \sum_{i,j=1}^{n} x_i y_i (\epsilon_i \epsilon_j - \epsilon_j \epsilon_i) \\ &= \sum_{i,j=1}^{n} x_i y_j \sum_m (r^j_{im} - r^i_{jm}) \epsilon_m . \end{aligned} \tag{57}$$

If L contains all $\alpha\beta - \beta\alpha$

$$\sum_m c_m \sum_{i,j} x_i y_j (r^j_{im} - r^i_{jm}) = 0.$$

But the x_i, y_j are arbitrary, so that

$$\sum_m r^j_{im} c_m = \sum_m r^i_{jm} c_m . \tag{58}$$

From (58) the matrix $T = (\sum_m r^j_{im} c_m)_{ij}$ is symmetric. Conversely, by a reversal of the steps, we may show that the existence of a symmetric matrix with determinant different from zero, which transforms $\mathfrak{R}$ into $\mathfrak{S}$ requires the algebra $\mathfrak{a}$ to be symmetric.

THEOREM 6. *A necessary and sufficient condition for an algebra $\mathfrak{a}$ to be symmetric is that there exists a symmetric matrix T with determinant different from zero which transforms $\mathfrak{R}$ into $\mathfrak{S}$.*

We can now show

THEOREM 7. *If the algebra $\mathfrak{a}$ is symmetric then the indecomposable constituent $\mathfrak{U}_\kappa$ of $\mathfrak{R}$ is equivalent to the indecomposable constituent $\mathfrak{V}_\kappa$ of $\mathfrak{S}$ ($\kappa = 1, 2, \cdots, k$).*[23]

[23] Cf. N. N.

PROOF. We suppose $\mathfrak{U}_\kappa = \mathfrak{B}_\kappa$ for all κ is not true, and choose κ such that $\mathfrak{U}_\kappa = \mathfrak{B}_\lambda$, $\lambda \neq \kappa$. From Theorem 5 we then have

$$\begin{aligned} I(\mathfrak{U}_\kappa, \mathfrak{F}_\kappa) &= 1, \\ I(\mathfrak{F}_\kappa, \mathfrak{U}_\kappa) &= I(\mathfrak{F}_\kappa, \mathfrak{B}_\lambda) = 0. \end{aligned} \tag{59}$$

Let us, therefore, examine matrices intertwining $\mathfrak{U}_\kappa = (u_{ij})$ with $\mathfrak{F}_\kappa = (f_{ij})$ and compare these with matrices intertwining $\mathfrak{F}_\kappa$ with $\mathfrak{U}_\kappa$.

From the method of Theorem 1 we have that the matrices $Q_r = (u^i_{tr})_{ij}$, $r = 1, 2, \cdots, u_\kappa$, form a complete linearly independent set of matrices intertwining $\mathfrak{U}_\kappa$ with $\mathfrak{S}$

$$\mathfrak{U}_\kappa Q_r = Q_r \mathfrak{S} \qquad (r = 1, 2, \cdots, u_\kappa). \tag{60}$$

Again from Theorem 1, the matrices $\overline{P}_s = (f^i_{sj})_{ij}$ form a complete linearly independent set of matrices intertwining $\mathfrak{R}$ with $\mathfrak{F}_\kappa$,

$$\mathfrak{R}\overline{P}_s = \overline{P}_s \mathfrak{F}_\kappa \qquad (s = 1, 2, \cdots, f_\kappa). \tag{61}$$

By Theorem 6, since $\mathfrak{a}$ is symmetric, there exists a symmetric matrix T with $|T| \neq 0$ such that

$$\mathfrak{R} = T\mathfrak{S}T^{-1}, \qquad T = T'. \tag{62}$$

From (60), (62) and (61) it follows that

$$\mathfrak{U}_\kappa Q_r T^{-1} \overline{P}_s = Q_r T^{-1} \overline{P}_s \mathfrak{F}_\kappa. \tag{63}$$

The matrices $Q_r T^{-1} \overline{P}_s$, $r = 1, 2, \cdots, u_\kappa$; $s = 1, 2, \cdots, f_\kappa$ intertwine $\mathfrak{U}_\kappa$ with $\mathfrak{F}_\kappa$.

From (39) the matrix $Z = \begin{pmatrix} 0 \\ E_{f_\kappa} \end{pmatrix}$ with u_κ rows and f_κ columns intertwines $\mathfrak{U}_\kappa$ with $\mathfrak{F}_\kappa$. We may arrange the basis elements (26) so that

$$\mathfrak{R} = \begin{pmatrix} \mathfrak{U}_\kappa & & & \\ & \mathfrak{U}_* & & 0 \\ & & \ddots & \\ & 0 & & \mathfrak{U}_* \end{pmatrix}. \tag{64}$$

Then

$$\mathfrak{U}_\kappa (E_{u_\kappa}, 0) = (E_{u_\kappa}, 0)\mathfrak{R},$$

which gives on setting

$$\begin{aligned} (E_{u_\kappa}, 0) &= X, \\ \mathfrak{U}_\kappa X T &= X T \mathfrak{S}. \end{aligned} \tag{65}$$

By (65) XT is a matrix intertwining $\mathfrak{U}_\kappa$ with $\mathfrak{S}$, and may, therefore, be written as a linear sum, say $\sum_r a_r Q_r$ of the matrices Q_r, so that

$$X = \sum_r a_r Q_r T^{-1}. \tag{66}$$

On the other hand from (39) and (64)

$$\mathfrak{R}Y = Y\mathfrak{F}_\kappa ,$$

where $Y = \begin{pmatrix} 0 \\ E_{f_\kappa} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ is a matrix of n rows, f_κ columns, having a unit matrix in the block of f_κ rows following the $(u_\kappa - f_\kappa)^{\text{th}}$ row. Since Y intertwines $\mathfrak{R}$ and $\mathfrak{F}_\kappa$ it may be written as a linear combination, say $\sum_s b_s \overline{P}_s$ of the matrices $\overline{P}_s$. It follows that

$$XY = (E_{u_\kappa}, 0) \begin{pmatrix} 0 \\ E_{f_\kappa} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ E_{f_\kappa} \end{pmatrix} = Z = \sum_{r,s} a_r b_s Q_r T^{-1} \overline{P}_s .$$

Consequently, the matrix Z which intertwines $\mathfrak{U}_\kappa$ with $\mathfrak{F}_\kappa$ is a linear combination of the matrices $Q_r T^{-1} \overline{P}_s$ and so at least one of these matrices is different from the zero matrix. Now if $T^{-1} = (\gamma_{ij})$, $Q_r T^{-1} \overline{P}_s = \left(\sum_{j,k} u^j_{ir} \gamma_{jk} f^k_{sl} \right)_{il}$ and for some choice of the subscripts i, r, s, l we must have a non-zero coefficient, say

$$\sum_{j,k} u^j_{wx} \gamma_{jk} f^k_{yz} \neq 0. \tag{67}$$

By a similar method we may find matrices intertwining $\mathfrak{F}_\kappa$ with $\mathfrak{U}_\kappa$. As in (60) we have

$$\mathfrak{F}_\kappa \bar{Q}_r = \bar{Q}_r \mathfrak{S} \qquad (r = 1, 2, \cdots, f_\kappa) \tag{68}$$

where $\bar{Q}_r = (f^i_{ir})_{ij}$. Again, as in (61)

$$\mathfrak{R}P_s = P_s \mathfrak{U}_\kappa \qquad (s = 1, 2, \cdots, u_\kappa) \tag{69}$$

where $P_s = (u^i_{sj})_{ij}$. Then from (68), (62), and (69) we obtain

$$\mathfrak{F}_\kappa \bar{Q}_r T^{-1} P_s = \bar{Q}_r T^{-1} P_s \mathfrak{U}_\kappa$$
$$(r = 1, 2, \cdots, f_\kappa ; s = 1, 2, \cdots, u_\kappa).$$

However, by (59) $I(\mathfrak{F}_\kappa , \mathfrak{U}_\kappa) = 0$, so

$$\bar{Q}_r T^{-1} P_s = \left(\sum_{j,k} f^j_{ir} \gamma_{jk} u^k_{sl} \right)_{il} = 0. \tag{70}$$

The coefficients of these matrices (70) are zero for all choices of subscripts i, r, s, l and, in particular, for $i = y, r = z, s = w, l = x$, that is,

$$\sum_{j,k} f^j_{yz} \gamma_{jk} u^k_{wx} = \sum_{j,k} f^k_{yz} \gamma_{kj} u^j_{wx} = 0. \tag{71}$$

But the matrix T is symmetric, and the same is, therefore, true for $T^{-1} = (\gamma_{jk})$, with the result that the sum in (71) is the same as that in (67), which provides a contradiction. Then $\mathfrak{U}_\kappa$ must be equivalent to $\mathfrak{V}_\kappa$ for $\kappa = 1, 2, \cdots, k$.

The equivalence of $\mathfrak{U}_\kappa$ and $\mathfrak{V}_\kappa$ in the case of a symmetric algebra has two immediate implications:

(1) By Theorems 2 and 2′ the first and the last irreducible constituent of $\mathfrak{U}_\kappa$ are both $\mathfrak{F}_\kappa$.

(2) $\mathfrak{F}_\kappa$ has the same multiplicity as irreducible constituent in $\mathfrak{V}_\lambda$ as in $\mathfrak{U}_\lambda$ so that from Theorems 4 and 4′ $c_{\kappa\lambda} = c_{\lambda\kappa}$.

THEOREM 8. *If the algebra $\mathfrak{a}$ is symmetric, then the first and the last irreducible constituent of the indecomposable constituent $\mathfrak{U}_\kappa$ of $\mathfrak{R}$ are the same. Moreover $c_{\kappa\lambda} = c_{\lambda\kappa}$ for all κ, λ.*

The semisimple algebras in algebraically closed fields are symmetric. Let $\mathfrak{a}$ be such an algebra, and $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ a basis of $\mathfrak{a}$. The (reduced) discriminant matrix $D = (S(\epsilon_i\epsilon_j))_{ij}$ (where S is the reduced trace of $\mathfrak{a}$, that is, the sum of the traces of the distinct irreducible representations of $\mathfrak{a}$ and $S(\epsilon)$ is the value of this trace for the element ϵ of $\mathfrak{a}$) has determinant different from zero.[24] Further from the linearity of S we have $D = \left(\sum_m r^j_{im} S(\epsilon_m)\right)_{ij}$ so that by Theorem 1, Cor. 1, D intertwines $\mathfrak{R}$ and $\mathfrak{S}$ and hence transforms $\mathfrak{R}$ into $\mathfrak{S}$. Evidently D is symmetric, so from Theorem 6 the statement follows.

Again let $G_1, G_2, \cdots, G_g$ be the elements of a finite group $\mathfrak{G}$ of order g and suppose $G_1 = E$ the unit element. Associated with $\mathfrak{G}$ we have the group ring Γ consisting of all elements $\gamma = x_1G_1 + x_2G_2 + \cdots + x_gG_g$ where the x_i are elements from a field K. Γ is an algebra over K, semisimple in the case that the characteristic of the field K is relatively prime to g, but not semisimple if the characteristic divides g.

From equations of the form (1) and (2), where the basis elements ϵ_i are now the elements G_i of $\mathfrak{G}$ we obtain the regular representations $\mathfrak{R}$ and $\mathfrak{S}$ of Γ. In particular, we have

$$G_iG_j = \sum_k r^j_{ik} G_k = \sum_k \delta_{ij,k} G_k$$

where

$$\delta_{ij,k} = 1 \quad \text{if} \quad G_iG_j = G_k,$$

$$\delta_{ij,k} = 0 \quad \text{if} \quad G_iG_j \neq G_k.$$

Let us consider the hyperplane L consisting of all elements $\sum_i x_iG_i$ of Γ with $x_1 = 0$. The commutator element $\alpha\beta - \beta\alpha$ of Γ where $\alpha = \sum_i a_iG_i$, $\beta = \sum_i b_iG_i$ may be expressed as

$$\alpha\beta - \beta\alpha = \sum_{i,j} a_ib_j(G_iG_j - G_jG_i) = \sum_{i,j} a_ib_j(\delta_{ij,1} - \delta_{ji,1})G_1 + \sum_{i,j} a_ib_j \sum_{k\neq 1} (\delta_{ij,k} + \delta_{ji,k})G_k.$$

[24] Cf. M. Deuring, *Algebren*, Ergebn. Math. 4 (1935) p. 33. And in this connection see also footnote 16 of N. N.

It is easily seen that $\delta_{ij,1} = \delta_{ji,1}$ so that $x_1 = 0$ contains $\alpha\beta - \beta\alpha$. Again, let us suppose $x_1 = 0$ contains the right ideal (ξ) where $\xi = \sum_i \mu_i G_i$. We have

$$\xi \cdot G_j = \sum_i \mu_i G_i G_j = \sum_i \mu_i \delta_{ij,1} G_1 + \sum_i \mu_i \sum_{k \neq 1} \delta_{ij,k} G_k$$
$$= \mu_m G_m + \sum_i \mu_i \sum_{k \neq 1} \delta_{ij,k} G_k,$$

where $G_m = G_j^{-1}$. Requiring that $\xi \cdot G_j$ lies in $x_1 = 0$ we obtain $\mu_m = 0$, and applying the argument for each G_j we see that $\mu_i = 0$ $(i = 1, 2, \cdots, g)$ i.e. (ξ) is the null ideal.

This proof is still valid in the case that the underlying field has for its characteristic a prime which divides the order of G so that Γ is not semisimple. As a particular consequence all the foregoing theorems are applicable to the representations of finite groups in Galois fields.[25]

[25] Cf. R. Brauer-C. Nesbitt, *On the Modular Representations of Groups of Finite Order*, Univ. Toronto Studies, Math. Series 4, (1937).

UNIVERSITY OF TORONTO AND THE INSTITUTE FOR ADVANCED STUDY.

ANNALS OF MATHEMATICS
Vol. 39, No. 3, July, 1938

NOTE ON SYMMETRIC ALGEBRAS[1]

BY TADASI NAKAYAMA AND CECIL NESBITT

(Received March 21, 1938)

In studying the regular representations of algebras R. Brauer and one of the writers of the present note introduced recently the notion of symmetric algebras;[2] an algebra is called symmetric if there exists a hyperplane in the algebra which contains all commutator elements $ab - ba$, but does not contain any right ideal (except the null ideal). Let $F_1, F_2, \cdots, F_k$ be the distinct irreducible representations of an algebra A, and $U_1, U_2, \cdots, U_k (V_1, V_2, \cdots, V_k)$ the corresponding indecomposable parts of the first (second) regular representation of A.[3] The main theorem on symmetric algebras states that for a symmetric algebra U_κ is equivalent to $V_\kappa (\kappa = 1, 2, \cdots, k)$. However an algebra for which $U_\kappa \cong V_\kappa$ (for all κ) is not necessarily symmetric, and hence it may be desirable to obtain a condition which is not only sufficient but also necessary for the equivalence of U_κ and V_κ (for all κ), and which is an immediate weakening of that of symmetric algebras. The purpose of the present note is to give such a condition; we shall see that in order that U_κ and V_κ be equivalent (for all κ) it is both necessary and sufficient that in the algebra there exists a hyperplane which contains no right ideal (except the null ideal), but contains all elements of the form $e_i a e_j (i \neq j)$, where $e_1, e_2, \cdots, e_l$ are mutually orthogonal idempotent elements with the sum 1 such that the right ideals $e_i A$ are directly indecomposable. Furthermore we shall study some elementary problems concerning the behavior of symmetric algebras under change and extension of the underlying fields.

We are grateful to Prof. Brauer, many of whose ideas we have exploited in the following.[4]

1. Let K be a field and let A be an (associative) algebra (with a finite basis) over K. Let $a \to M(a)$ be a representation of A (in K) and let $\mathfrak{m}$ be the cor-

[1] Presented to the American Mathematical Society, April 15, 1938. The present formulation was done by Nakayama.

[2] R. Brauer and C. Nesbitt, *On the regular representations of algebras*, Proc. Nat. Acad. Sci. **23** (1937); for a complete account embodying Nesbitt's thesis (Univ. Toronto) see Ann. Math. **39** (1938), pp. 634; for the regular representations of algebras see also T. Nakayama, *Some studies on regular representations, induced representations and modular representations*, Ann. Math. **39** (1938), p. 361. We refer to these papers as B. N. Proc., N. Th., and N. Ann. respectively.

[3] See the three papers in the footnote 2.

[4] Cf. the forthcoming paper of R. Brauer.

responding left representation module[5] with a K-basis $(u_1, u_2, \cdots, u_m)$, $a(u_i) = (u_i)M(a)$; m is the degree of the representation $M(a)$. Consider, furthermore, a left ideal $\mathfrak{b}$ of A, and let $(b_1, b_2, \cdots, b_n)$ be a K-basis of $\mathfrak{b}$. Take any element w of $\mathfrak{m}$. Then the mapping $\theta: b \to b \cdot w$ is an (A-operator) homorphism of $\mathfrak{b}$ onto the submodule $\mathfrak{b} \cdot w$ of $\mathfrak{m}$. If $w = \sum u_i \mu_i = (u_i)(\mu_i)$, then $b \cdot w = (u_1, u_2, \cdots, u_m)M(b)(\mu_i)$, in particular,

$$(b_1, \cdots, b_n)w = (u_1, \cdots, u_m)(M(b_1)(\mu_i), \cdots, M(b_n)(\mu_i)).$$

Hence the rank of $\mathfrak{b} \cdot w$ with respect to K is equal to the rank of the (m, n) matrix[6]

$$P = (M(b_1)(\mu_i), \cdots, M(b_n)(\mu_i)). \tag{1}$$

Then the map group $\mathfrak{b}^\theta = \mathfrak{b} \cdot w$ coincides with the whole of $\mathfrak{m}$ if and only if P has the rank m.[7]

LEMMA 1. *Let $\mathfrak{c}$ be a sub-(left) ideal of $\mathfrak{b}$. If there exists an $(m, 1)$-matrix, that is, an m-dimensional vector, (μ_i) such that*

$$M(c)(\mu_i) = 0 \qquad \textit{for all } c \textit{ in } \mathfrak{c}, \tag{2}$$

and such that the matrix P (in (1)) has the rank m for this (μ_i), then the representation of A defined by the residue class module $\mathfrak{b}/\mathfrak{c}$, considered as a left representation module of A, can be reduced to the form

$$\begin{pmatrix} M(a) & 0 \\ * & * \end{pmatrix}. \tag{3}$$

PROOF. From the above consideration we know that the mapping $\theta: b \to b \cdot w$ with $w = (u_i)(\mu_i)$ is a homomorphism of $\mathfrak{b}$ onto the whole of $\mathfrak{m}$. Moreover, (2) implies $\mathfrak{c}^\theta = 0$. Hence θ induces a homomorphism of $\mathfrak{b}/\mathfrak{c}$ on the whole of $\mathfrak{m}$.

---

[5] We shall call a representation module of A a left or right representation module according as A is considered as its left or its right operator domain. For the fundamental concepts in the representation theory see for instance, E. Noether, *Hyperkomplexe Grössen und Darstellungstheorie*, Math. Zeits. **30** (1929); B. L. van der Waerden, *Gruppen von linearen Transformationen*, Ergebn. Math. **4** (1935); H. Weyl, *Gruppentheorie und Quantenmechanik*, Leipzig (1928).

[6] In the special case where $\mathfrak{b} = A$ and $M(a)$ is the first regular representation of A defined by the basis $(b_1, \cdots, b_n)$ of $\mathfrak{b} = A$ the matrix P becomes the so-called *parastrophic* matrix of A. See G. Frobenius, *Theorie der hyperkomplexen Grössen*, Sitz. Ber. Berlin, (1903).

[7] If this is the case it follows also that $\mathfrak{b}$ has an element e_0 such that $e_0^2 \equiv e_0 \pmod{\mathfrak{c}_0}$ and $(be_0, \mathfrak{c}_0) = (Ae_0, \mathfrak{c}_0) = \mathfrak{b}$, where $\mathfrak{c}_0$ is the subideal of $\mathfrak{b}$ consisting of all elements mapped onto 0 (the zero of $\mathfrak{m}$) by θ, whence $\mathfrak{b}/\mathfrak{c}_0 \cong \mathfrak{m}$. For $\mathfrak{b} \cdot w = \mathfrak{m}$ implies the existence of an element e_0 such that $e_0 w = w$. Then $(e_0^2 - e_0)w = 0$ whence $e_0^2 \equiv e_0 \pmod{\mathfrak{c}_0}$, and obviously $(be_0, \mathfrak{c}_0) = (Ae_0, \mathfrak{c}_0) = \mathfrak{b}$.

In particular, if the parastrophic determinant of an algebra A does not vanish identically, then we know that A has a right unit element 1_R. The (right, left)-dual consideration shows the existence of a left unit element 1_L under the same condition, and then $1 = 1_R = 1_L$ is a (unique) unit element of A. This is one of the two main theorems of Frobenius concerning the parastrophic matrix of an algebra. The purpose of the present note is to refine the other of them.

Then $\mathfrak{p} = \mathfrak{b}/\mathfrak{c}$ has a submodule $\mathfrak{p}_0$ such that the residue class module $\mathfrak{p}/\mathfrak{p}_0$ is isomorphic to $\mathfrak{m}$. This proves our lemma.

LEMMA 2. *Let $\mathfrak{b} = Ae$ be a left ideal of A generated by an idempotent element e in A, and let $\mathfrak{c}$ be a subideal of $\mathfrak{b}$. Then the existence of a vector (μ_i) satisfying the conditions in Lemma 1 is (not only sufficient but also) necessary in order that the representation defined by $\mathfrak{b}/\mathfrak{c}$ can be reduced in the form* (3).

PROOF. In this special case every homomorphism of $\mathfrak{b}$ into $\mathfrak{m}$ can be obtained in the manner described above. For, if θ is any homomorphism of $\mathfrak{b}$ into $\mathfrak{m}$ and $e^\theta = w(\epsilon\, \mathfrak{m})$ then θ is given by $b \to b^\theta = (be)^\theta = be^\theta = b \cdot w$. Hence the lemma is immediate.

REMARK 1. Let $\bar{b}_1, \bar{b}_2, \cdots, \bar{b}_n$ denote the residue classes of $b_1, b_2, \cdots, b_n$ (mod $\mathfrak{c}$) and let U be a matrix such that $(\bar{b}_1, \bar{b}_2, \cdots, \bar{b}_n)U$ is a K-basis of $\mathfrak{b}/\mathfrak{c}$. (The number of rows in U is n and the number of columns in U is the rank of $\mathfrak{b}/\mathfrak{c}$ with respect to K). Then the matrix $PU = (M(b_1)(\mu_i), \cdots, M(b_n)(\mu_i))U$ with (μ_i) satisfying (2) intertwines[8] $M(a)$ with the representation defined by our basis $(\bar{b}_1, \bar{b}_2, \cdots, \bar{b}_n)U$ of $\mathfrak{b}/\mathfrak{c}$. Moreover, if $\mathfrak{b}$ is, as in Lemma 2, a left ideal generated by an idempotent element, then every matrix intertwining $M(a)$ with the representation defined by $(\bar{b}_1, \bar{b}_2, \cdots, \bar{b}_n)U$ can be expressed in the above form (with a suitable (μ_i) satisfying (2)). These follow easily from the above considerations and the well known correspondence between homomorphisms and intertwining matrices.

Let us next consider a *right* representation module $\mathfrak{n}$ corresponding to our representation $M(a)$, and let (v_i) be its K-basis such that $(v_i)a = M(a)(v_i)$. Looking upon $\mathfrak{n}$ as an m-dimensional vector space, consider a hyperplane H in $\mathfrak{n}$ consisting of all elements

$$x = \xi_1 v_1 + \xi_2 v_2 + \cdots + \xi_m v_m = (\xi_i)(v_i) \text{ with}$$

$$\xi_1 \mu_1 + \xi_2 \mu_2 + \cdots + \xi_m \mu_m = (\xi_i)(\mu_i) = 0.$$

LEMMA 3. *The condition* (2) *is equivalent to that the hyperplane H contains $\mathfrak{n} \cdot \mathfrak{c}$. The matrix P* (in (1)) *has the rank m if and only if there is no non-zero element x in $\mathfrak{n}$ such that $x \cdot \mathfrak{b} \subseteq H$.*

PROOF. The first statement is obvious, for $(v_i)c = M(c)(v_i)$. To prove the second one let $x = (\xi_i)(v_i)$ be an element in $\mathfrak{n}$. That $x \cdot \mathfrak{b} \subseteq H$, that is, $x \cdot b = (\xi_i)(v_i)b = (\xi_i)M(b)(v_i) \in H$ for all b in $\mathfrak{b}$ means that $(\xi_i)M(b)(\mu_i) = 0$ for all b in $\mathfrak{b}$, and this is in turn equivalent to $(\xi_i)P = 0$. Hence, if P has the rank m, $x \cdot \mathfrak{b} \subseteq H$ implies $(\xi_i) = 0$, that is, $x = 0$. If, on the contrary, P has a rank less than m, we take a non-trivial solution (ξ_i) of $(\xi_i)P = 0$ and put $x = (\xi_i)(v_i)$. Then $x \cdot \mathfrak{b} \subseteq H$. These prove our lemma.

From these lemmas we obtain immediately

THEOREM 1. *Let e be an idempotent element in A, and let $\mathfrak{c}$ be a sub-(left) ideal*

[8] For the definition of a matrix which intertwines two representations see, for instance, B. N. Proc.

of the left ideal Ae generated by e. Let $a \to M(a)$ be a representation of A, and $\mathfrak{n}$ the corresponding right representation module. Then the representation of A defined by the left representation module $Ae/\mathfrak{c}$ can be reduced to the form (3) *if and only if there exists a hyperplane H in $\mathfrak{n}$ which contains $\mathfrak{n}\cdot\mathfrak{c}$ and such that there is no non-zero element x in $\mathfrak{n}$ satisfying $x\cdot Ae \subseteq H$.*[9]

2. Let A be again an algebra over K. We assume in the following that A has a unit element 1. Let N be the radical of A. A can be decomposed into a direct sum of directly indecomposable left ideals Ae_i generated by mutually orthogonal idempotent elements e_i with the sum $\sum e_i = 1$,

$$A = Ae_1 + Ae_2 + \cdots + Ae_l .$$

Then

$$A = e_1A + e_2A + \cdots + e_lA$$

is a decomposition of A into a direct sum of directly indecomposable right ideals. Let U_i and V_i be the representations of A defined by the right ideal e_iA and the left ideal Ae_i respectively. They are respectively indecomposable constituents of the first and the second regular representations of A. The last (first irreducible constituent[10] of $U_i(V_i)$ is unique and is defined by $e_iA/e_iN(Ae_i/Ne_i)$. Here the irreducible representations defined by e_iA/e_iN and Ae_i/Ne_i coincide, or are equivalent to each other, and we denote them by F_i. If F_i is equivalent to $F_j(F_i \cong F_j)$, then $U_i \cong U_j$ and $V_i \cong V_j$. Hence we have the one-to-one correspondence between distinct irreducible representations and the distinct indecomposable constituents of the first, as well as of the second, regular representation.[11] A is called *Frobeniusean* if the first and second regular representations are equivalent.[12] As was shown in N. Th. by an example, the indecomposable constituents U_i and V_i of the first and second regular representations corresponding to the same irreducible representation F_i are in general not equivalent to each other, even if A is Frobeniusean. In the following we shall examine the properties of A for which $U_i \cong V_i$ for $i = 1, 2, \cdots, l$. Let us call such an algebra A *weakly symmetric*. We prove

THEOREM 2. *In order that A be weakly symmetric, it is necessary and sufficient that in A, considered as a vector space, there exists a hyperplane H which contains e_iAe_j for all i, j with $i \neq j$, but does not contain any right ideal of A (except the null ideal).*

---

[9] If A has a unit element 1, then in the special case where $e = 1$ and $M(a)$ is, as usual, a representation whose matrix corresponding to 1 is the unit matrix, the last condition in Theorem 1 is equivalent to that H contains no $Ae(=A)$-right ideal.

[10] The reduction shall always be done in such a manner that the right upper part is 0.

[11] For all this see B. N. Proc., N. Th. and N. Ann. The notation here is the same as that of N. Ann.

[12] In order that A is Frobeniusean, it is necessary and sufficient that the parastrophic determinant of A does not identically vanish.

PROOF. i) Let H be a hyperplane in A having the above properties. The intersection $H \frown e_iA = e_iH$ is a hyperplane in e_iA (since $H \not\supseteq e_iA$) and does not contain any right ideal of A. Moreover, $H \frown e_iA \supseteq \sum_{j\neq i} e_iAe_j = e_iA \sum_{j\neq i} Ae_j$. Hence,[13] applying Theorem 1 to $e = 1$, $\mathfrak{c} = \sum_{j\neq i} Ae_j$ and $\mathfrak{n} = e_iA$, we find that the representation V_i defined by $Ae_i \cong A/\mathfrak{c}$ can be reduced to the form $\begin{pmatrix} U_i & 0 \\ * & * \end{pmatrix}$. Since this is the case for all i, we have $U_i \cong V_i$.

ii) Conversely, let A be weakly symmetric. Apply Theorem 1 again to $e = 1$, $\mathfrak{c} = \sum_{j\neq i} Ae_j$ and $\mathfrak{n} = e_iA$. It follows that e_iA has a hyperplane H_i containing $e_iA \sum_{j\neq i} Ae_j = \sum_{j\neq i} e_iAe_j$ and containing no A-right ideal. Let $v_{i\rho}(\rho = 1, 2, \cdots, m_i)$ be a K-basis of e_iA and let the hyperplane H_i be given by the equation $\sum_\rho \xi_{i\rho}\mu_{i\rho} = 0$ where $\xi_{i\rho}$ are the coördinates with respect to $v_{i\rho}$. Let then H be the hyperplane in A consisting of all elements

$$x = \sum_i \sum_\rho \xi_{i\rho} v_{i\rho} \quad \text{with} \quad \sum_i \sum_\rho \xi_{i\rho}\mu_{i\rho} = 0.$$

We have $H \frown e_iA = H_i$ and $H \supseteq e_iAe_j$ for all $i \neq j$. Furthermore, H contains no right ideal of A. For, if $\mathfrak{r}$ is a right ideal of A contained in H, then $e_i\mathfrak{r} \subseteq H \frown e_iA = H_i$, whence $e_i\mathfrak{r} = 0$. Since this is valid for every i, we have $\mathfrak{r} = 0$. These prove our theorem.

REMARK 2. The condition in Theorem 2 may be replaced by the slightly weaker one that there exists a hyperplane in A which contains all e_iAe_j, with $F_i \not\cong F_j (i, j = 1, 2, \cdots l)$, but does not contain any right ideal of A. One can see this easily by applying Theorem 1 to $e = 1$, $\mathfrak{c} = \sum_{F_j \not\cong F_i} Ae_j$ and $\mathfrak{n} = e_iA$. (Observe that only F_j which are equivalent to F_i can appear as first irreducible constituent of the representation defined by $A/\mathfrak{c}$. Further the existence of H containing no right ideal secures for us that A is Frobeniusean). The sum of these e_iAe_j with $F_i \not\cong F_j$ forms a module $\mathfrak{q}$ lying in the radical, and it is well known[14] that A can be expressed as the direct sum of $\mathfrak{q}$ and of primary algebras. Then our condition reads: There is a hyperplane in A which contains the module $\mathfrak{q}$ but does not contain any right ideal of A.

Let $i_\kappa(\kappa = 1, 2, \cdots, k)$ be a subset of the i's which correspond to distinct irreducible representations. We may then again weaken our condition to the form that there is a hyperplane in A which contains all $e_{i_\kappa}Ae_j$ with $F_j \not\cong F_{i_\kappa}(\kappa = 1, 2, \cdots, k; j = 1, 2, \cdots, l)$. This follows immediately by a similar application of Theorem 1.

Furthermore, if we apply Theorem 1 to $e = e_i$, $\mathfrak{c} = 0$, and $\mathfrak{n} = e_iA$, then we find that in order that A be weakly symmetric it is also necessary and sufficient that there is a hyperplane H in A such that for every $i = 1, 2, \cdots, l$ (or, merely

[13] Cf. the footnote 9.

[14] Cf. M. Deuring, *Algebren*, Ergebn. Math. **4** (1935), p. 17.

for $i = i_\kappa(\kappa = 1, 2, \cdots, k)$) there is no non-zero element x in e_iA satisfying $x \cdot Ae_i \subsetneq H \frown e_iH$.

REMARK 3. In connection with the above considerations the following fact may be of some interest. *Let U be an indecomposable representation of an algebra A which can be defined by a left ideal Ae as well as by a right ideal $e'A$, where e and e' are idempotent elements of A. Then, if a representation of A has U as its constituent, U is a direct constituent of it;*[15] *in particular, U can not be a constituent of another indecomposable representation.*

To prove this, let $a \to M(a)$ be a representation of A having U as its constituent,

$$M(a) \cong \begin{pmatrix} * & 0 & 0 \\ * & U & 0 \\ * & * & * \end{pmatrix}$$

Let $\mathfrak{m}$ and $\mathfrak{n}$ be the left and right representation moduli of $M(a)$ respectively. Then $\mathfrak{m}$ has two submoduli $\mathfrak{m}_0$ and $\mathfrak{m}_1$ such that $\mathfrak{m}_0/\mathfrak{m}_1$ is isomorphic to Ae. Let u be an element of $\mathfrak{m}_0$ whose residue class (mod $\mathfrak{m}_1$) corresponds to e by this isomorphism. Then Aeu $(= Au)$ must be isomorphic to Ae. Since $\mathfrak{m}$ has a submodule Aeu isomorphic to Ae, the representation $M(a)$ can be reduced in the form $\begin{pmatrix} * & 0 \\ * & U \end{pmatrix}$. But this means the existence of a submodule $\mathfrak{n}_0$ of $\mathfrak{n}$ such that $\mathfrak{n}/\mathfrak{n}_0 \cong e'A$. Let v be an element of $\mathfrak{n}$ whose residue class (mod $\mathfrak{n}_0$) corresponds to e' by the isomorphism. Then $ve'A$ $(= vA) \cong e'A$ and $\mathfrak{n} = (ve'A, \mathfrak{n}_0)$. Considering the ranks of $\mathfrak{n}$, $\mathfrak{n}_0$ and $ve'A$, we see that $\mathfrak{n}$ is the direct sum of $ve'A$ and $\mathfrak{n}_0$. Hence U is a direct constituent of $M(a)$.

Now, we define, as in B. N. Proc. and N. Th., a *symmetric* algebra in the following manner: A is called symmetric, if there is a hyperplane in A which contains all commutator elements $ab - ba$ $(a, b \in A)$ but does not contain any right ideal (except the null ideal). This condition is equivalent to the existence of a non-singular symmetric matrix which transforms the first regular representation into the second regular representation (defined by the same basis of A).[16]

Since each element e_iae_j in e_iAe_j $(i \neq j)$ is a commutator $e_iae_j = e_iae_j - ae_je_i$, we have

THEOREM 3. *If A is symmetric, then A is weakly symmetric, that is, $U_i \cong V_i$ for all i.*

An example of an algebra which is weakly symmetric, but is not symmetric,

---

[15] R. Brauer has proved a related lemma.

[16] For instance, semi-simple algebras and group rings (of finite groups over any field) are symmetric. To prove that semi-simple algebras are symmetric, it may be used that their (reduced) discriminants are not zero (cf. N. Th.). This proof fails in the case where the underlying field is not perfect (=*vollkommen*). But this case can be reduced to the case of a perfect underlying field by means of Theorem 5 below.

is the following:[17] Let A be the algebra of rank 4 with the basis elements

$$a_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \alpha & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

$$a_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \beta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}; \quad a_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where α, β are two different non-zero elements in the underlying field K: $0 \neq \alpha \neq \beta \neq 0$. Then the parastrophic matrix of the algebra A is

$$\begin{pmatrix} 0 & 0 & 0 & \mu_1 \\ 0 & 0 & \alpha\mu_1 & \mu_2 \\ 0 & \beta\mu_1 & 0 & \mu_3 \\ \mu_1 & \mu_2 & \mu_3 & \mu_4 \end{pmatrix}$$

and this has the determinant 0 as soon as it is symmetric. Hence A is not symmetric, but is obviously weakly symmetric.

REMARK 4. In Theorem 2 and 3 we considered only right ideals. But the hyperplanes there do not contain any left ideal either. In fact, a hyperplane in A contains no right ideal if and only if it contains no left ideal. This is immediate since the criterion can be reduced in both cases to the non-vanishing of the parastrophic determinant when for the values of the parameters are taken the coefficients which define the hyperplane.

3. Let K_0 be a subfield of our underlying field K such that K is of a finite degree, say h, over K_0. Then the algebra A can be considered also as an algebra over K_0. To distinguish this, let us denote A by A/K or A/K_0 according as it is considered as an algebra over K or K_0.

THEOREM 4. *A/K_0 is symmetric, (or weakly symmetric, or Frobeniusean) if and only if A/K is symmetric (or weakly symmetric, or Frobeniusean).*

PROOF. Let $\kappa_1, \kappa_2, \cdots, \kappa_h$ be a K_0-basis of K, and let $\kappa \to S(\kappa)$ be the second regular representation of K/K_0 with respect to the basis (κ_σ). Then the transposed representation $S'(\kappa)$ is the first regular representation of K/K_0 with respect to the same basis. K/K_0 is obviously symmetric, hence there exists a non-singular symmetric matrix T with elements from K_0 such that $TS(\kappa) = S'(\kappa)T$ (for all κ in K).

LEMMA 4. *Let $\mathfrak{m}$ be a left representation module of A/K with a K-basis $(u_1,$*

[17] Another example will be given in §3. An essential difference between the examples is that the algebra A here remains weakly symmetric after any extension of the underlying field, but the same does not hold for the A in §3.

$u_2, \cdots, u_m$) *defining a representation* $a \to M(a) = (\alpha_{ij}(a))_{ij}$ *of* A/K. *Then* $\mathfrak{m}$ *can be considered as a representation module of* A/K_0 *with a* K_0*-basis* $(u_{i\kappa_\sigma})$ $(i = 1, 2, \cdots, m; \sigma = 1, 2, \cdots, h)$ *and in this sense it defines the representation* $a \to (S(\alpha_{ij}(a)))_{ij}$ *of* A/K_0 (*of degree* mh). *If our representation module* $\mathfrak{m}$ *is a right module, then we have only to replace* S *by* S'.

PROOF. This can be verified easily.

a) Now let A/K be symmetric. Let $a \to N(a) = (\beta_{ij}(a))$ and $a \to M(a) = (\alpha_{ij}(a))$ be the first and second regular representations of A/K respectively. Then the above lemma shows that $a \to (S'(\beta_{ij}(a)))$ and $a \to (S(\alpha_{ij}(a)))$ are the first and second regular representations of A/K_0. In virtue of our assumption there exists a non-singular symmetric matrix $P = (\pi_{ij})$ with elements from K transforming $N(a)$ into $M(a)$. Then

$$(S(\pi_{ij}))(S(\alpha_{ij}(a))) = (S(\beta_{ij}(a)))(S(\pi_{ij})),$$

$$(S'(\pi_{ij}))(S'(\alpha_{ij}(a))) = (S'(\beta_{ij}(a)))(S'(\pi_{ij})).$$

Hence

$$(E \times T)(S(\pi_{ij}))(S(\alpha_{ij}(a))) = (S'(\pi_{ij}))(S'(\alpha_{ij}(a)))(E \times T)$$

$$= (S'(\beta_{ij}(a)))(S'(\pi_{ij}))(E \times T) = (S'(\beta_{ij}(a)))(E \times T)(S(\pi_{ij})),$$

where E is the unit matrix of order $r = (A:K)$ and $E \times T$ is the Kronecker product of E and T. The matrix $(E \times T)(S(\pi_{ij}))$ is obviously non-singular, and is symmetric too, for

$$((E \times T)(S(\pi_{ij})))' = (S(\pi_{ij}))'(E \times T)' = (S'(\pi_{ji}))(E \times T')$$

$$= (S'(\pi_{ij}))(E \times T) = (E \times T)(S(\pi_{ij})).$$

Therefore, A/K_0 is symmetric.

b) Conversely let A/K_0 be symmetric. Then there exists a non-singular symmetric matrix U with elements from K_0 such that

$$U(S(\alpha_{ij}(a))) = (S'(\beta_{ij}(a)))U,$$

whence

$$(E \times T)^{-1}U(S(\alpha_{ij}(a))) = (S(\beta_{ij}(a)))(E \times T)^{-1}U. \tag{7}$$

LEMMA 5. $(E \times T)^{-1}U$ *can be expressed in the form* $(E \times T)^{-1}U = (S(\pi_{ij}))_{ij}$ *for suitable (and uniquely determined)* π_{ij} *in* K.

PROOF. Let $(E \times T)^{-1}U$ be written in the form $(V_{ij})_{ij}$ where the V_{ij} are square matrices of order h. Put $a = \kappa\ (\epsilon K)$ in (7). Since $\alpha_{ij}(\kappa) = \beta_{ij}(\kappa) = \kappa\delta_{ij}$ (Kronecker δ), it follows from (7) that every V_{ij} commutes with $S(\kappa)$. Now, the set $(K_0)_h$ of all square matrices of order h with elements from K_0 form a normal simple algebra over K_0, and $S(\kappa)(\kappa \in K)$ form a subfield of $(K_0)_h$ with degree h over K_0. It is well known[18] then that the field formed of the $S(\kappa)$ is a

[18] See, for example, R. Brauer, *Über die algebraische Struktur von Schiefkörpern*, Crelles Journal, **166** (1932); E. Noether, *Nichtkommutative Algebra*, Math. Zeits. **37** (1933).

maximal commutative subsystem of $(K_0)_h$, that is, every element in $(K_0)_h$ which commutes with all $S(\kappa)$ has the form $S(\pi)(\pi \in K)$. Hence $V_{ij} = S(\pi_{ij})$ for suitable π_{ij}. The π_{ij} are uniquely determined since the representation $\kappa \to S(\kappa)$ is faithful.

Now put $P = (\pi_{ij})$. Then, converting the whole process in i, a), one can easily see that P is a non-singular symmetric matrix transforming $N(a) = (\beta_{ij}(a))$ into $M(a) = (\alpha_{ij}(a))$. Thus A/K is symmetric.

ii) That A/K_0 is Frobeniusean if and only if A/K is so, can be seen in quite a similar way as in i). We have only to drop the symmetry of P, U (and of T if we wish).

iii) If we apply Lemma 4 to Ae_i and e_iA, then it is also easy to see that A/K_0 is weakly symmetric if and only if A/K is so.

Let us next go to another problem which is concerned also with the behavior of symmetric algebras with respect to underlying fields. Let L be any extension field of K, that is, a field containing K, and let A_L denote the algebra over L obtained from $A = A/K$ by extending the underlying field to L.

THEOREM 5. *If A is symmetric, then A_L is obviously so. The converse is also true, provided that the field K contains more elements than the rank $r = (A:K)$ of A.*[19] *A_L is Frobeniusean if and only if A is so. If A_L is weakly symmetric, then A is so, but the converse is in general not true.*

PROOF. The first statement is obvious. To prove the second one, let $a_1, a_2, \cdots, a_r$ be a K-basis of A, and let $N(a)$ and $M(a)$ be the first and second regular representations of A/K. Consider a system of linear equations and an inequality

$$x_{ij} = x_{ji},\ (x_{ij})M(a_p) = N(a_p)(x_{ij}) \qquad (p = 1, 2, \cdots, r);\ \det |x_{ij}| \neq 0$$

for r^2 indeterminates x_{ij}. If A_L is symmetric, then this system has a solution in L. Then it is well known that the system has also a solution in K, provided that K has more elements than the degree of the inequality, that is, r.

The statement concerning Frobeniusean algebras, whose first half is trivial, follows immediately from the well known theorem that two representations which are equivalent in an extension field are already equivalent in the original field; this theorem can be proved in a similar way as above in the case where the original field has sufficiently many elements, but the theorem is, by an argument of R. Brauer and E. Noether, valid also in the contrary case.[20] The first statement concerning weakly symmetric algebras can be proved easily also by the same theorem. That, even if A is weakly symmetric, A_L is not necessarily so (due to the fact that an irreducible representation in K may be reducible into different irreducible parts in L) can be seen from the following example:

EXAMPLE. Let K be a field such that there exists a separable quadratic

[19] We can not now say whether this restriction is material or not.

[20] See, for instance, B. L. van der Waerden, *Gruppen von linearen Transformationen*, Ergebn. Math. **4** (1935) p. 70.

extension field L. Consider the algebra of rank 4 over K consisting of all matrices

$$\begin{pmatrix} \alpha & & & \\ \beta & \alpha' & & \\ & & \alpha' & \\ & & \beta' & \alpha \end{pmatrix}; \alpha, \beta \in L,$$

where α', β' denote the conjugates of α, β respectively. A is weakly symmetric, as one can easily see. But A_L is formed of all matrices

$$\begin{pmatrix} \alpha & & & \\ \beta & \alpha_1 & & \\ & & \alpha_1 & \\ & & \beta_1 & \alpha \end{pmatrix}; \alpha, \alpha_1, \beta, \beta_1 \in L$$

and A_L is not weakly symmetric.[21]

INSTITUTE FOR ADVANCED STUDY.

[21] A is also an example of an algebra which is weakly symmetric but is not symmetric. Cf. footnote 17.

ANNALS OF MATHEMATICS
Vol. 39, No. 3, July, 1938

ON CYCLIC ALGEBRAS[1]

BY A. ADRIAN ALBERT

(Received March 14, 1938)

1. Introduction

Let an algebra $\mathfrak{Z}$ of order n over a field $\mathfrak{K}$ be a direct sum of a finite number of equivalent cyclic fields $\mathfrak{C}_i$. Then $\mathfrak{Z}$ has an automorphism S of order n over $\mathfrak{K}$. Such algebras are in a sense generalizations of cyclic fields and it is natural to define a corresponding generalization of cyclic algebras. It turns out that the resulting algebras are normal simple algebras similar to cyclic algebras.

Put $\mathfrak{Z}$, S and $\mathfrak{Z}'$, S' into the same class $\mathfrak{A} = (\mathfrak{Z}, S)$ if $\mathfrak{Z}$ and $\mathfrak{Z}'$ are equivalent with respect to a simple isomorphism

$$z \leftrightarrow z', \qquad (z \text{ in } \mathfrak{Z}, z' \text{ in } \mathfrak{Z}')$$

such that $(z^S)' = z'^{S'}$. We shall call these classes *cyclic systems*. O. Teichmüller[2] considered the set $\sum$ of all such classes defined by algebras $\mathfrak{Z}$ of the same order n and defined an operation on $\sum\sum$ to $\sum$. By the use of certain properties of cyclic algebras he proved that $\sum$ is an abelian group with respect to this operation and that the order of each $(\mathfrak{Z}, S)$ is the degree of the corresponding cyclic fields $\mathfrak{C}_i$.

It is possible to prove the properties of $\sum$ *independently of the theory of cyclic algebras* and we shall give these proofs here. They provide a very elegant basis for the corresponding theorems on cyclic algebras. In particular consider the case where $\mathfrak{Z}$ is a cyclic field of degree n over $\mathfrak{K}$, $n = qd$, $\mathfrak{Y}$ is the subfield of degree q over $\mathfrak{K}$ of $\mathfrak{Z}$. Then the author proved[3] in 1932 that if $(\mathfrak{Z}, S, \gamma)$ is a corresponding cyclic algebra we have

$$(\mathfrak{Z}, S, \gamma)^d \sim (\mathfrak{Y}, S, \gamma).$$

This result *was used* by Teichmüller to obtain a corresponding property of $(\mathfrak{Z}, S)^d$. However while the known proof of the cyclic algebra result is quite

[1] Presented to the Society April 8, 1938.

[2] Deutsche Mathematik, vol. 1 (1936), pp. 197–238. He called the algebras $\mathfrak{Z}$ cyclic normal rings and, while he stated his theorems in terms of such normal rings, he had previously defined equivalence in the sense given above. Thus his results really do not concern $\mathfrak{Z}$ but rather the cyclic system $(\mathfrak{Z}, S)$.

[3] In the paper on the construction of cyclic algebras with a given exponent, American Journal of Mathematics, vol. 54 (1932), pp. 1–13. In that paper I first reduced the question to the case $n = p^s$, p a prime, and then gave the proof for this case. Later Hasse stated and proved the same result for the general n giving essentially the same proof (Math. Ann., vol. 107 (1933), pp. 731–760). The result has important applications. In particular see the application to theorems on norms in the American Journal paper.

complicated and not too natural *we shall obtain a very simple proof of the corresponding theorem on cyclic systems.* Our proofs thus provide an elegant new proof of the important cyclic algebra theorem stated above.

2. Algebras with automorphisms

We shall begin our investigations by discussing some elementary properties[4] of algebras with automorphisms. All of our concepts will be defined relative to a fixed coefficient field $\mathfrak{K}$. Consider an algebra $\mathfrak{Z}$ of finite order over $\mathfrak{K}$ and let $\mathfrak{Z}$ have a unity element which we designate by 1. Then we call $\mathfrak{Z}$ a diagonal algebra if $\mathfrak{Z} = (e_1, \cdots, e_t)$ over $\mathfrak{K}$ for pairwise orthogonal idempotents e_i. Thus a diagonal algebra over $\mathfrak{K}$ is equivalent to the set of all t-rowed diagonal matrices with elements in $\mathfrak{K}$, and $e_1 + \cdots + e_t = 1$ is then the identity matrix. A simple computation shows that *the only idempotent elements of a diagonal algebra are sums of a finite number of distinct e_i.*

An automorphism S over $\mathfrak{K}$ of an algebra $\mathfrak{Z}$ carries any subalgebra $\mathfrak{B}$ of $\mathfrak{Z}$ into an equivalent subalgebra $\mathfrak{B}^S$. When $\mathfrak{B} = \mathfrak{B}^S$ the correspondence $b \leftrightarrow b^S$ defines an automorphism $b \leftrightarrow b^S$ of $\mathfrak{B}$. We make the

DEFINITION. *Let $\mathfrak{G}$ be a group of automorphisms over $\mathfrak{K}$ of $\mathfrak{Z}$. Then $\mathfrak{Z}$ will be called $\mathfrak{G}$-reducible if $\mathfrak{Z}$ is the direct sum $\mathfrak{B} \oplus \mathfrak{C}$ with $\mathfrak{B} = \mathfrak{B}^S$, $\mathfrak{C} = \mathfrak{C}^S$ for every S of $\mathfrak{G}$. Otherwise $\mathfrak{Z}$ is called $\mathfrak{G}$-irreducible.*

We use the terminology S-reducible and S-irreducible when $\mathfrak{G} = [S]$ is the cyclic group generated by S. Observe that if I is the identity automorphism the concept of I-irreducibility coincides with the classical one of irreducibility.[5] We now prove

THEOREM 1. *Every $\mathfrak{G}$-reducible algebra $\mathfrak{Z}$ is uniquely expressible, apart from the order of its components, as a direct sum of $\mathfrak{G}$-irreducible algebras.*[6]

The result is well known[7] for the case of the identity group and we may write $\mathfrak{Z} = \mathfrak{C}_1 \oplus \cdots \oplus \mathfrak{C}_g$ with $\mathfrak{C}_i$ irreducible. Then also $\mathfrak{Z} = \mathfrak{C}_1^S \oplus \cdots \oplus \mathfrak{C}_g^S$ where $\mathfrak{C}_i^S$ is equivalent to $\mathfrak{C}_i$ and hence is irreducible for every S of $\mathfrak{G}$. But then the $\mathfrak{C}_i^S$ coincide with the $\mathfrak{C}_i$ in some order. If $\mathfrak{C}$ is one of the $\mathfrak{C}_i$ the algebras $\mathfrak{C}, \mathfrak{C}^{S_2}, \cdots, \mathfrak{C}^{S_t}$ are distinct for a finite set of S_i of $\mathfrak{G}$ and $\mathfrak{C}^S = \mathfrak{C}^{S_j}$ for every S of $\mathfrak{G}$. Hence $(\mathfrak{C}^{S_j})^S = \mathfrak{C}^{S_k}$ for every S of $\mathfrak{G}$, $\mathfrak{C}_i^S = \mathfrak{C}^{S_j}$ if and only if $\mathfrak{C}_i$ is one of the $\mathfrak{C}^{S_k}$. Clearly

$$\mathfrak{B}_1 = \mathfrak{C}_1 \oplus \mathfrak{C}_1^{S_2} \oplus \cdots \oplus \mathfrak{C}_1^{S_t} \tag{1}$$

is a $\mathfrak{G}$-irreducible component of $\mathfrak{Z}$ and $\mathfrak{Z} = \mathfrak{B}_1 \oplus \cdots \oplus \mathfrak{B}_s$ for unique $\mathfrak{G}$-irreducible $\mathfrak{B}_i$ all of the form $\mathfrak{C}_{j_i} \oplus \cdots \oplus \mathfrak{C}_{j_i}^{S_g}$.

---

[4] While most of these properties are essentially contained in Teichmüller's paper our exposition would lose in elegance if we picked isolated results from that paper. Thus we shall make our discussion entirely independent of his.

[5] For the concept of irreducibility see L. E. Dickson, Algebren und ihre Zahlentheorie, p. 86.

[6] Our result is probably deducible from considerations of groups with operators but is most simply a consequence of the well known case $\mathfrak{G} = [I]$.

[7] Dickson, p. 88.

THEOREM 2. *Every $\mathfrak{G}$-irreducible algebra is expressible as the direct product $\mathfrak{Z} = \mathfrak{C} \times \mathfrak{E}$ with $\mathfrak{C}$ irreducible and $\mathfrak{E}$ a uniquely determined diagonal algebra. The algebra $\mathfrak{C}$ is unique in the sense of equivalence.*

For by the proof of Theorem 1 we may write $\mathfrak{Z} = \mathfrak{C}_1 \oplus \mathfrak{C}_1^{s_2} \oplus \cdots + \mathfrak{C}_1^{s_t}$ with $\mathfrak{C}_1$ irreducible. We let c_1 range over all quantities of $\mathfrak{C}_1$ and write

$$c = c_1 + c_1^{s_2} + \cdots + c_1^{s_t}. \tag{2}$$

Then the algebra $\mathfrak{C}$ of all quantities c is equivalent to $\mathfrak{C}_1$ and thus is irreducible. If e_1 is the unity element of $\mathfrak{C}_1$, that of $\mathfrak{Z}$ and of $\mathfrak{C}$ is $1 = e_1 + e_1^{s_2} + \cdots + e_1^{s_t}$. We write $\mathfrak{E} = (e_1, \cdots, e_1^{s_t})$ over $\mathfrak{K}$ and an elementary argument shows that $\mathfrak{Z} = \mathfrak{C} \times \mathfrak{E}$. If also $\mathfrak{Z} = \mathfrak{C}_0 \times \mathfrak{E}_0$ where $\mathfrak{C}_0$ is irreducible and $\mathfrak{E}_0 = (f_1, \cdots, f_h)$ is a diagonal algebra, the idempotents f_j are unity elements of the irreducible components $\mathfrak{C}_0 f_j$ of $\mathfrak{Z}$ and thus the f_j must be equal to the $e_1^{s_j}$, the $\mathfrak{C}_0 f_j$ equal to the $\mathfrak{C}_1^{s_i}$, $\mathfrak{E} = \mathfrak{E}_0$. Finally the algebras $\mathfrak{C}_0$, $\mathfrak{C}_1 = \mathfrak{C}_0 f_{j_1} = \mathfrak{C}e_1$, and $\mathfrak{C}$ are all equivalent.[8]

Observe that the algebras $\mathfrak{C}$ and $\mathfrak{E}$ of our result have the properties $\mathfrak{C} = \mathfrak{C}^{s^t}$, $\mathfrak{E} = \mathfrak{E}^S$ for every S of $\mathfrak{G}$. Thus every S generates an automorphism S_0 of $\mathfrak{C}$ and we may prove

THEOREM 3. *Let $\mathfrak{Z}$ be S-irreducible so that $\mathfrak{Z} = \mathfrak{C} \times \mathfrak{E}$, where $\mathfrak{C}$ is irreducible and $\mathfrak{E}$ is a diagonal algebra of order t. Then $\mathfrak{C}$ has an automorphism of order s and the order of S in $\mathfrak{Z}$ is st. Conversely let $\mathfrak{Z} = \mathfrak{C} \times \mathfrak{E}$ where $\mathfrak{E}$ is diagonal of order t and $\mathfrak{C}$ is irreducible with an automorphism of order s. Then $\mathfrak{Z}$ is S-irreducible with respect to an automorphism S of order st.*

For let $\mathfrak{Z} = \mathfrak{C}_1 \oplus \mathfrak{C}_1^S \oplus \cdots \oplus \mathfrak{C}_1^{s^{t-1}}$ with $\mathfrak{C}_1 = \mathfrak{C}e_1$ by the proof of Theorem 2. The general quantity of $\mathfrak{C}$ is $c = c_1 + c_1^S + \cdots + c_1^{S^{r-1}}$ and if $T = S^t$ the correspondence $c \leftrightarrow c^T$ defines an automorphism T_0 of $\mathfrak{C}$ whose order is that of the automorphism T_1 induced in $\mathfrak{C}_1$ by T. But the order of S is then clearly st. Conversely let $\mathfrak{Z} = \mathfrak{C} \times \mathfrak{E}$ so that if $\mathfrak{E} = (e_1, \cdots, e_1^t)$ the general quantity of $\mathfrak{Z}$ is

$$a = c_1 e_1 + c_2 e_2 + \cdots + c_t e_t \qquad (c_j \text{ in } \mathfrak{C}). \tag{3}$$

Then if T is an automorphism of order s of $\mathfrak{C}$ we write

$$a^S = c_1 e_2 + c_2 e_3 + \cdots + c_{t-1} e_t + c_t^T e_1 \tag{4}$$

and see that $a \leftrightarrow a^S$ is an automorphism S of $\mathfrak{Z}$. But $a^{S^t} = c_t^T e_1 + \cdots + c_t^T e_t$, S has order st in $\mathfrak{Z}$. The assumption that $\mathfrak{C}$ is irreducible combined with $\mathfrak{Z} = \mathfrak{C}_1 \oplus \mathfrak{C}_1^S \oplus \cdots \oplus \mathfrak{C}_1^{s^{t-1}}$, $\mathfrak{C}_1 = \mathfrak{C}e_1$, $\mathfrak{C}_1^{S^i} = \mathfrak{C}e_{i+1}$, implies that $\mathfrak{Z}$ is S-irreducible. The order of an automorphism S of an S-reducible algebra $\mathfrak{Z}$ is clearly the

[8] However while $\mathfrak{E} = \mathfrak{E}_0$ we may take $\mathfrak{C} \neq \mathfrak{C}_0$. For example let $\mathfrak{K}$ have characteristic not two, a be a non-square of $\mathfrak{K}$, $\mathfrak{C}_1 = (e_1, i)$ over $\mathfrak{K}$, $i^2 = ae_1$. Also let S have order 2, $\mathfrak{A} = \mathfrak{C}_1 + \mathfrak{C}_1^S$. Then $\mathfrak{A} = \mathfrak{C} \times \mathfrak{E} = \mathfrak{C}_0 \times \mathfrak{E}_0$ where $\mathfrak{C} = (1, i + i^S)$ over $\mathfrak{K}$, $\mathfrak{C}_0 = (1, i - i^S)$ over $\mathfrak{K}$, $\mathfrak{C} \neq \mathfrak{C}_0$.

least common multiple of its orders in the S-irreducible components of $\mathfrak{Z}$. We shall be particularly interested in a type of algebra $\mathfrak{Z}$ of order n over $\mathfrak{K}$ with an automorphism S of order n over $\mathfrak{K}$. This assumption may easily be shown *not* to have the consequence that $\mathfrak{Z}$ is S-irreducible. However we shall add additional restrictions which imply the truth of this latter conclusion.

3. Semi-fields

Let $\mathfrak{Z}$ be a semi-simple algebra over $\mathfrak{K}$. Then we say that $\mathfrak{Z}$ is *separable* (over $\mathfrak{K}$) if the scalar extension $\mathfrak{Z}_\mathfrak{L}$ is semi-simple for every scalar extension field $\mathfrak{L}$ over $\mathfrak{K}$. It is well known[9] that irreducible components of $\mathfrak{Z}$ are simple algebras $\mathfrak{Z}_i$ each with a centrum $\mathfrak{K}_i$ over $\mathfrak{K}$ and that $\mathfrak{Z}$ is separable if and only if the $\mathfrak{K}_i$ are all separable fields over $\mathfrak{K}$.

If $\mathfrak{Z}$ is a commutative semi-simple algebra its irreducible components are all algebraic fields over $\mathfrak{K}$ and we shall call $\mathfrak{Z}$ a *semi-field*. A separable semi-field $\mathfrak{Z}$ over $\mathfrak{K}$ has the property that $\mathfrak{Z}_\mathfrak{L}$ is a diagonal algebra for an algebraic extension $\mathfrak{L}$ of finite degree over $\mathfrak{K}$.

DEFINITION. *Let $\mathfrak{Z}$ be a separable semi-field of order n over $\mathfrak{K}$ and let S be an automorphism of order n. Then we call $\mathfrak{Z}$ a cyclic semi-field with generating automorphism S if $\mathfrak{Z}_\mathfrak{L} = (e, e^S, \cdots e^{S^{n-1}})$ over $\mathfrak{L}$ for a field $\mathfrak{L}$ over $\mathfrak{K}$ and pairwise orthogonal idempotents e^{S^i}.*

Diagonal algebras are trivially cyclic semi-fields. Also cyclic fields are known[10] to be cyclic semi-fields. In fact we may prove the consequent

THEOREM 4. *A semi-field $\mathfrak{Z}$ of order n over $\mathfrak{K}$ is cyclic if and only if it is the direct product of a cyclic field $\mathfrak{C}$ of degree s over $\mathfrak{K}$ and a diagonal algebra $\mathfrak{E}$ of order t over $\mathfrak{K}$ such that $n = st$.*

For let $\mathfrak{Z}$ be a cyclic semi-field. If $\mathfrak{Z}$ were S-reducible then $\mathfrak{Z}_\mathfrak{L}$ is also S-reducible with $\mathfrak{Z}_\mathfrak{L} = (e, e^S, \cdots, e^{S^{n-1}})$. This is contrary to Theorem 2. Hence $\mathfrak{Z} = \mathfrak{C} \times \mathfrak{E}$ where $\mathfrak{C} = \mathfrak{C}^{S^t}$ and Theorem 3 states that the order of S is the order t of $\mathfrak{E}$ over $\mathfrak{K}$ multiplied by the order of S^t in $\mathfrak{C}$. But the order of S is equal to the order of $\mathfrak{Z}$ over $\mathfrak{K}$ and $\mathfrak{Z} = \mathfrak{C} \times \mathfrak{E}$. Hence $\mathfrak{C}$ is a cyclic field with generating automorphism S^t. Conversely let $\mathfrak{Z} = \mathfrak{C} \times \mathfrak{E}$ where $\mathfrak{E} = (e_1, \cdots, e_t)$ over $\mathfrak{K}$ and $\mathfrak{C}$ is cyclic of degree s with generating automorphism T. The general quantity of $\mathfrak{Z}$ is $z = c_1e_1 + \cdots + c_te_t$ with c_i in $\mathfrak{C}$ and the automorphism

$$z \leftrightarrow z^S = c_1e_2 + \cdots + c_{t-1}e_t + c_t^Te_1$$

is an automorphism of order n of $\mathfrak{Z}$. There exists a field $\mathfrak{L}$ over $\mathfrak{K}$ such that $\mathfrak{C}_\mathfrak{L} = (e, e^T, \cdots, e^{T^{n-1}})$ and a simple computation shows that $\mathfrak{Z}_\mathfrak{L}$ is the diagonal algebra with basal elements the idempotents

$$e^{T^{i-1}}e_k = (ee_1)^{S^{(i-1)t+(k-1)}}. \tag{5}$$

The proof above has the immediate consequences

COROLLARY I. *A semi-field $\mathfrak{Z}$ of order n over $\mathfrak{K}$ with an automorphism S of*

[9] Cf. M. Deuring, *Algebren*, p. 17 and pp. 36ff.

[10] Cf. the author's *Modern Higher Algebra*, p. 240, equation (61).

order n is a cyclic semi-field with generating automorphism S if and only if $\mathfrak{Z}$ is S-irreducible.

COROLLARY II. *The direct product of two diagonal algebras is a diagonal algebra.*

We also have

COROLLARY III. *Let $\mathfrak{Z}$ be a cyclic semi-field over $\mathfrak{K}$ and $\mathfrak{L}$ be any scalar extension field of $\mathfrak{K}$. Then $\mathfrak{Z}_{\mathfrak{L}}$ is a cyclic semi-field over $\mathfrak{L}$.*

COROLLARY IV. *The only quantities of a cyclic semi-field unaltered by its generating automorphism S are the quantities of $\mathfrak{K}$.*

If $\mathfrak{X}$ is a subalgebra of $\mathfrak{Z}$ with the same unity element as $\mathfrak{Z}$ and such that $\mathfrak{X}$ is a field over $\mathfrak{K}$ we call $\mathfrak{X}$ a sub-field of $\mathfrak{Z}$. Then $\mathfrak{X} = \mathfrak{K}(x)$ where $x = c_1e_1 + \cdots + c_te_t$ for the c_i and e_i as above. Clearly $f(x) = 0$ if and only if $f(c_i) = 0$, that is $\mathfrak{X}$ is equivalent to the sub-fields $\mathfrak{K}(c_i)$ of $\mathfrak{C}$. Moreover $\mathfrak{C}$ is a subfield of $\mathfrak{Z}$ of maximal degree and we have

COROLLARY V. *The sub-fields of a cyclic semi-field $\mathfrak{Z} = \mathfrak{C} \times \mathfrak{E}$ are all equivalent to sub-fields of $\mathfrak{C}$. Moreover $\mathfrak{Z}$ has a largest sub-field $\mathfrak{C}$ which is unique in the sense of equivalence.*

DEFINITION. *The cyclic semi-fields $\mathfrak{Z}_1$ and $\mathfrak{Z}_2$ are called similar and we write $\mathfrak{Z}_1 \sim \mathfrak{Z}_2$ if their largest sub-fields are equivalent.*

Note that all diagonal algebras over $\mathfrak{K}$ are similar to $\mathfrak{K}$.

4. Direct products

We shall consider direct products of cyclic semi-fields and shall use

THEOREM 5. *A cyclic semi-field $\mathfrak{Z}$ with generating automorphism S has a normal basis,*

$$\mathfrak{Z} = (u, u^S, \cdots, u^{S^{n-1}}) \text{ over } \mathfrak{K}. \tag{6}$$

For $\mathfrak{Z} = \mathfrak{C} \times \mathfrak{E}$ where $\mathfrak{C}$ is a cyclic field with generating automorphism G, $\mathfrak{E} = (e_1, \cdots, e_t)$ for pairwise orthogonal idempotents e_i. The general quantity of $\mathfrak{Z}$ has the form

$$z = c_1e_1 + \cdots + c_te_t \qquad (c_i \text{ in } \mathfrak{C}), \tag{7}$$

and the generating automorphism S is the correspondence $z \leftrightarrow z^S$ where

$$z^S = c_1e_2 + c_2e_3 + \cdots + c_{t-1}e_t + c_t^Ge_1. \tag{8}$$

It is known[11] that $\mathfrak{C}$ has a normal basis $v, v^G, \cdots, v^{G^{s-1}}$ where s is the degree of $\mathfrak{C}$ over $\mathfrak{K}$. Then $\mathfrak{Z}$ has a basis

$$v^{G^i}e_j \qquad (i = 0, \cdots, s-1; j = 1, \cdots, t). \tag{9}$$

Put $u = ve_1$ and see that $u^S = ve_2$, $u^{S^2} = ve_3$, $\cdots$, $u^{S^t} = v^Ge_1$,

$$u^{S^q} = v^{G^i}e_{j+1} \qquad (q = it + j;\ i = 0, \cdots, s-1;\ j = 0, \cdots, t-1). \tag{10}$$

[11] Ibid, p. 201.

Hence (6) holds.

Let $\mathfrak{Z} = \mathfrak{Z} \times \mathfrak{Y}$ where $\mathfrak{Z}$ and $\mathfrak{Y}$ are cyclic semi-fields of the same order n. We let $z_1, z_2, \cdots, z_n$ be any basis of $\mathfrak{Z}$ over $\mathfrak{K}$. Then the general quantity of $\mathfrak{Z}$ has the form

$$a = z_1 b_1 + \cdots + z_n b_n \qquad (b_i \text{ in } \mathfrak{Y}). \tag{11}$$

Define

$$a^{S^i} = z_1^{S^i} b_1 + \cdots + z_n^{S^i} b_n, \tag{12}$$

$$a^{T^j} = z_1 b_1^{T^j} + \cdots + z_n b_n^{T^j}. \tag{13}$$

Then the correspondences $a \leftrightarrow a^{S^i}$, $a \leftrightarrow a^{T^j}$ are automorphisms of $\mathfrak{Z}$ over $\mathfrak{K}$ such that the first induces the automorphism S^i in $\mathfrak{Z}$ and leaves the quantities of $\mathfrak{Y}$ unaltered, the second induces the automorphism T^j in $\mathfrak{Y}$ and leaves the quantities of $\mathfrak{Z}$ unaltered. We designate these automorphisms of $\mathfrak{Z}$ by S^i and T^j respectively and note that we are using the notation of the automorphisms of the individual factors $\mathfrak{Z}$ and $\mathfrak{Y}$ for the induced automorphisms of $\mathfrak{Z}$. With this convention we prove

THEOREM 6. *Let $\mathfrak{W}_k$ be the set of all quantities of $\mathfrak{Z} = \mathfrak{Z} \times \mathfrak{Y}$ unaltered by ST^k. Then ST^k induces the generating automorphism S of $\mathfrak{Z}$, $\mathfrak{W}_k$ is a cyclic semi-field of order n over $\mathfrak{K}$ with generating automorphism induced by T, and*

$$\mathfrak{Z} = \mathfrak{Z} \times \mathfrak{W}_k. \tag{14}$$

For proof take the z_i of (11) to be a normal basis of $\mathfrak{Z}$ over $\mathfrak{K}$, $z_i = z_1^{S^i}$. Then

$$a^{ST^k} = z_2 b_1^{T^k} + \cdots + z_n b_{n-1}^{T^k} + z_1 b_n^{T^k} = a$$

if and only if

$$b_1 = b_n^{T^k}, \qquad b_2 = b_1^{T^k}, \cdots, \qquad b_n = b_{n-1}^{T^k}. \tag{15}$$

The last $n - 1$ of these equations determine $b_2, \cdots, b_n$ uniquely in terms of b_1. The first implies only that $b_1 = b_1^{T^{nk}}$, a consequence of our assumption that T^n is the identity automorphism of $\mathfrak{Z}$. Hence a is in $\mathfrak{W}_k$ for any element b_1 of $\mathfrak{Y}$ providing that $b_2, \cdots, b_n$ are then determined by (15). It follows that $\mathfrak{W}_k$ is a linear set of order n over $\mathfrak{K}$ and in fact that if $y_1, \cdots, y_n$ are a normal basis of $\mathfrak{Y}$ over $\mathfrak{K}$ then

$$w_i^{(k)} = y_i z_1 + y_i^{T^k} z_2 + \cdots + y_i^{T^{(n-1)k}} z_n \tag{16}$$

form a basis of $\mathfrak{W}_k$ over $\mathfrak{K}$. The defining property of $\mathfrak{W}_k$ clearly implies that $\mathfrak{W}_k$ is an algebra over $\mathfrak{K}$. Moreover $(w_i^{(k)})^T = w_{i+1}^{(k)}$ so that T induces an automorphism of order n in $\mathfrak{W}_k$. The unity elements of $\mathfrak{Z}$, $\mathfrak{Y}$, $\mathfrak{Z}$ and $\mathfrak{W}_k$ all coincide and if $\mathfrak{W}_k$ were T-reducible there would be an idempotent $e \neq 1$ in $\mathfrak{W}_k$ such that $e = e^T$. Hence $e = e^{T^{-k}}$, $e = e^{ST^k}$, $e = e^S$. The only elements a of $\mathfrak{Z}$ such that $a = a^S = a^T$ are elements of $\mathfrak{K}$ so that e is in $\mathfrak{K}$, $e = 1$, a contradiction. Hence $\mathfrak{W}_k$ is a cyclic semi-field with (16) as normal basis.

To prove (14) we observe that the product linear set $\mathfrak{Z}\mathfrak{W}_k$ is unaltered by every S^iT^j since $\mathfrak{Z}^{S^iT^j} = \mathfrak{Z}^{S^i} = \mathfrak{Z}$, $(\mathfrak{W}_k)^{S^iT^j} = \mathfrak{W}_k^{T^{j-ki}} = \mathfrak{W}_k$. There exists a field $\mathfrak{L}$ over $\mathfrak{K}$ such that $\mathfrak{Z}_\mathfrak{L}$ and $\mathfrak{Y}_\mathfrak{L}$ are diagonal. Diagonal algebras have normal bases composed of idempotents and we let e generate such a basis of $\mathfrak{Z}_\mathfrak{L}$, f of $\mathfrak{Y}_\mathfrak{L}$. By our proof and (16) the algebra $(\mathfrak{W}_k)_\mathfrak{L}$ has a basis

$$g_i^{(k)} = f_ie_1 + f_{i+k}e_2 + \cdots + f_{i+(n-1)k}e_n . \tag{17}$$

Then $(\mathfrak{W}_k\mathfrak{Z})_\mathfrak{L}$ contains $g_1^{(k)}e_1 = f_ie_1$ and hence also $(f_ie_1)^{S^j} = f_ie_j$. But the f_ie_j form a basis of $\mathfrak{Z}_\mathfrak{L}$, $(\mathfrak{W}_k\mathfrak{Z})_\mathfrak{L}$ has order n^2 over $\mathfrak{L}$. Hence $\mathfrak{W}_k\mathfrak{Z}$ has order n^2 over $\mathfrak{K}$ and must be $\mathfrak{Z}$, $\mathfrak{Z} = \mathfrak{Z} \times \mathfrak{W}_k$. We have proved our theorem.

If $\mathfrak{Z} = \mathfrak{C} \times \mathfrak{E}$ where $\mathfrak{C}$ is a cyclic field and $\mathfrak{E}$ is a diagonal algebra we have seen that $\mathfrak{E}$ and $\mathfrak{C}$ may be chosen so that $\mathfrak{E} = (e_1, \cdots, e_t)$ with $e_i^S = e_{i+1}$, $\mathfrak{C}$ has generating automorphism S^t. We now prove

THEOREM 7. *Let $\mathfrak{Z} = \mathfrak{C} \times \mathfrak{E}$, $\mathfrak{Y} = \mathfrak{D} \times \mathfrak{E}_0$, where $\mathfrak{E}_0$ and $\mathfrak{E}$ have the same order t. Then if $\mathfrak{C}_k$ is the sub-algebra of $\mathfrak{C} \times \mathfrak{D}$ unaltered by $(ST^k)^t$ the algebra $\mathfrak{W}_k = \mathfrak{E}_k \times \mathfrak{C}_k$ with $\mathfrak{E}_k$ diagonal of order t.*

For $\mathfrak{Z} \times \mathfrak{Y} = \mathfrak{C} \times \mathfrak{E} \times \mathfrak{D} \times \mathfrak{E}_0$. Since S is a generating automorphism of $\mathfrak{E}$ and T of $\mathfrak{E}_0$ the subalgebra $\mathfrak{E}_k$ of $\mathfrak{E} \times \mathfrak{E}_0$ unaltered by ST^k has order t and is a subalgebra of $\mathfrak{W}_k$. The subalgebra of $\mathfrak{C} \times \mathfrak{D}$ unaltered by ST^k contains the subalgebra unaltered by $(ST^k)^t$ which is $\mathfrak{C}_k$ by Theorem 6. Hence $\mathfrak{C} \times \mathfrak{D} = \mathfrak{C} \times \mathfrak{C}_k$, $\mathfrak{E} \times \mathfrak{E}_0 = \mathfrak{E} \times \mathfrak{E}_k$, $\mathfrak{Z} \times \mathfrak{Y} = \mathfrak{C} \times \mathfrak{E} \times \mathfrak{C}_k \times \mathfrak{E}_k = \mathfrak{Z} \times (\mathfrak{C}_k \times \mathfrak{E}_k)$. But $\mathfrak{C}_k \times \mathfrak{E}_k$ has order n and is contained in $\mathfrak{W}_k$ of order n, $\mathfrak{C}_k \times \mathfrak{E}_k = \mathfrak{W}_k$.

5. Cyclic systems

The algebras $\mathfrak{W}_k$ depend not only upon $\mathfrak{Z}$ and $\mathfrak{Y}$ but also upon S and T. To express this dependence we make the

DEFINITION. *Let $\mathfrak{Z}$ and $\mathfrak{Z}'$ be cyclic semi-fields of order n with respective generating automorphisms S, S'. Then the pairs $\mathfrak{Z}'$, S' and $\mathfrak{Z}$, S are called equivalent if there exists a simple isomorphism*

$$z \leftrightarrow z' \qquad (z \text{ in } \mathfrak{Z}, z' \text{ in } \mathfrak{Z}')$$

such that

$$(z^S)' = (z')^{S'} \tag{18}$$

for every z of $\mathfrak{Z}$.

Thus $\mathfrak{Z}$, S and $\mathfrak{Z}'$, S' are called equivalent if the generating automorphisms S and S' induce each other under a simple isomorphism. It is trivial to see that when $\mathfrak{Z}$ and $\mathfrak{Z}'$ are diagonal of the same order the pairs $\mathfrak{Z}$, S and $\mathfrak{Z}'$, S' are always equivalent. In fact we have the following simple result whose proof is obvious.

THEOREM 8. *Two pairs $\mathfrak{Z}$, S and $\mathfrak{Z}'$, S' are equivalent if and only if the corresponding pairs $\mathfrak{C}$, S^t and $\mathfrak{C}'$, S'^t are equivalent.*

We put all pairs equivalent to $\mathfrak{Z}$, S into a class $(\mathfrak{Z}, S)$ which we shall call a

cyclic system of degree n. Then $(\mathfrak{Z}, S) = (\mathfrak{Z}', S')$ if and only if $(\mathfrak{C}, S^t) = (\mathfrak{C}', S'^t)$. Also if $(\mathfrak{Z}, S) = (\mathfrak{Z}', S')$ and $(\mathfrak{Y}, T) = (\mathfrak{Y}', T')$ the corresponding systems $(\mathfrak{W}_k, T)$ and $(\mathfrak{W}'_k, T')$ of Theorem 6 are equal. We now prove

THEOREM 9. *The set* $\sum$ *of all cyclic systems of the same degree* n *forms an abelian group with respect to the operation*[12]

$$(\mathfrak{Z}, S)\cdot(\mathfrak{Y}, T) = (\mathfrak{W}_{-1}, T). \tag{19}$$

The order in $\sum$ *of each system* $(\mathfrak{Z}, S)$ *is the degree of a largest subfield of* $\mathfrak{Z}$, *and the identity element of* $\sum$ *is* $\mathfrak{J} = (\mathfrak{E}, S)$ *where* $\mathfrak{E}$ *is diagonal. Finally*

$$(\mathfrak{W}_{-k}, T) = (\mathfrak{Z}, S)^k\cdot(\mathfrak{Y}, T). \tag{20}$$

For (19) defines a unique product system. The set of elements of $\mathfrak{Z} \times \mathfrak{Y}$ unaltered by ST^{-1} is the same as that unaltered by $TS^{-1} = (ST^{-1})^{-1}$ so that $(\mathfrak{Z}, S)\cdot(\mathfrak{Y}, T) = (\mathfrak{Y}, T)\cdot(\mathfrak{Z}, S)$. To prove the associative law we let $\mathfrak{V}$ be a cyclic semi-field of order n with generating automorphism U, and prove the

LEMMA. *Let* $\mathfrak{W}$ *be the set of all quantities of* $\mathfrak{Z} \times \mathfrak{Y}$ *unaltered by* ST^{-1}, $\mathfrak{W}_0$ *be the set of all quantities of* $\mathfrak{W} \times \mathfrak{V}$ *unaltered by* TU^{-1}, *so that*

$$\mathfrak{A} = (\mathfrak{Z} \times \mathfrak{Y}) \times \mathfrak{V} = (\mathfrak{Z} \times \mathfrak{W}) \times \mathfrak{V} = \mathfrak{Z} \times (\mathfrak{W} \times \mathfrak{V}) = \mathfrak{Z} \times (\mathfrak{W} \times \mathfrak{W}_0).$$

Then $\mathfrak{W}_0$ *is the set* $\mathfrak{X}$ *of all quantities of* $\mathfrak{A}$ *unaltered by both* ST^{-1} *and* TU^{-1}.

For in $(\mathfrak{Z} \times \mathfrak{W}) \times \mathfrak{V}$ all quantities of both $\mathfrak{W}$ and $\mathfrak{V}$ are unaltered by ST^{-1}, in $\mathfrak{W} \times \mathfrak{V}$ all quantities of $\mathfrak{W}_0$ are unaltered by TU^{-1}, $\mathfrak{W}_0 \leqq \mathfrak{X}$. Every quantity of $\mathfrak{A}$ has the unique form $a = \sum_{i=1}^{n} a_i u_i$ with the u_i a basis of $\mathfrak{W}_0$ over $\mathfrak{K}$ and the a_i in $\mathfrak{Z} \times \mathfrak{W}$. Then $a = a^{ST^{-1}} = a^{TU^{-1}}$ if and only if the $a_i = a_i^{ST^{-1}} = a_i^{TU^{-1}}$. However $z^U = z$, $w^U = w$ for every z of $\mathfrak{Z}$, w of $\mathfrak{W}$, so that our conditions imply that $a_i = a_i^T = a_i^{ST^{-1}}$, $a_i = a_i^S = a_i^T$, the a_i in $\mathfrak{Z} \times \mathfrak{W}$ must be in $\mathfrak{K}$. Thus a is in $\mathfrak{W}_0$, $\mathfrak{X} = \mathfrak{W}_0$.

The cyclic system $(\mathfrak{W}_0, U)$ is clearly the product $[(\mathfrak{Z}, S)\cdot(\mathfrak{Y}, T)]\cdot(\mathfrak{V}, U)$. By symmetry $(\mathfrak{Z}, S)\cdot[(\mathfrak{Y}, T)\cdot(\mathfrak{V}, U)] = (\mathfrak{W}_{10}, U)$, where $\mathfrak{W}_{10}$ is the set of all quantities of $\mathfrak{A}$ unaltered by TU^{-1} and SU^{-1}. However we have $ST^{-1} = (SU^{-1})(TU^{-1})^{-1}$ and $\mathfrak{W}_{10} = \mathfrak{W}_0$, $(\mathfrak{W}_{10}, U) = (\mathfrak{W}_0, U)$. This proves the associative law.

The direct product $\mathfrak{Z} \times \mathfrak{Y} = \mathfrak{Z} \times \mathfrak{W}_{-k}$ where ST^{-k} is a generating automorphism of $\mathfrak{Z}$, $\mathfrak{W}_{-k}$ has the generating automorphism T. Now $(\mathfrak{W}_{-1}, T) = (\mathfrak{Z}, S)\cdot(\mathfrak{Y}, T)$. Assume that $(\mathfrak{W}_{-(k-1)}, T) = (\mathfrak{Z}, S)^{k-1}\cdot(\mathfrak{Y}, T)$. Then $(\mathfrak{Z}, S) = (\mathfrak{Z}, ST^{-(k-1)})$, $\mathfrak{Z} \times \mathfrak{W}_{-(k-1)} = \mathfrak{Z} \times \mathfrak{W}_{-k}$ where $(\mathfrak{Z}, S)\cdot(\mathfrak{W}_{-(k-1)}, T) = (\mathfrak{W}_{-k}, T)$ since $ST^{-(k-1)}T^{-1} = ST^{-k}$. We have proved that (20) holds.

---

[12] We call $(\mathfrak{W}_{-1}, T)$ the *product* of $(\mathfrak{Z}, S)$ and $(\mathfrak{Y}, T)$. It is with Theorem 9 that our theory diverges from that of Teichmüller. This is really possible because of our Theorem 6 which is stronger than the corresponding partial result with $k = -1$ of Teichmüller. The reader will see, of course, that our proof does not depend at all upon the theory of cyclic algebras.

We complete our proof by the use of the following simple device. Take $\mathfrak{E} = (f_1, \cdots, f_n)$ over $\mathfrak{K}$. Here the f_i are pairwise orthogonal idempotent elements, we define $f_t = f_r$ for all integers $t = qn + r$, $0 < r \leqq n$ and have $f_i = f_{i-1}^T$. Then the algebra $\mathfrak{W}_{-k}$ has a basis

$$(w_1, \cdots, w_n) \text{ over } \mathfrak{K} \qquad w_i = w_1^{T^{i-1}} \qquad (i = 1, \cdots, n), \tag{21}$$

where by (16) with k replaced by $-k$ we have

$$w_1 = z_1 f_1 + z_2 f_{1-k} + \cdots + z_n f_{1-k(n-1)}. \tag{22}$$

In particular $w_1 = z_1 f_1 + z_2 f_n + \cdots + z_n f_2$ for $k = 1$ and $w_1^T = z_2 f_1 + z_3 f_n + \cdots + z_1 f_2$, $w_i^T = w_{i+1} = w_i^S$. The correspondence induced by $z_i \leftrightarrow w_i$ is a simple isomorphism of $\mathfrak{W}_{-1}$ and $\mathfrak{Z}$ preserving S and $(\mathfrak{W}_{-1}, T) = (\mathfrak{Z}, S)$. We have proved that $\mathfrak{J} = (\mathfrak{E}, T)$ is the identity element of $\sum$.

Now by (20) $(\mathfrak{W}_{-k}, T) = (\mathfrak{Z}, S)^k \cdot (\mathfrak{E}, T) = (\mathfrak{Z}, S)^k$. It follows that $(\mathfrak{Z}, S)^n = (\mathfrak{W}_{-n}, T)$. Since $ST^n = S$ and $\mathfrak{E}$ is the set of all quantities of $\mathfrak{Z}$ unaltered by S we have $(\mathfrak{Z}, S)^n = (\mathfrak{E}, T)$, $(\mathfrak{Z}, S)^{n-1}$ is the inverse in $\sum$ of $(\mathfrak{Z}, S)$, $\sum$ is a group. It remains to compute the order of the elements of $\sum$.

By Theorem 7 the algebra $\mathfrak{W}_{-k} = \mathfrak{E}_{-k} \times \mathfrak{C}_{-k}$ which is diagonal if and only if $\mathfrak{C}_{-k}$ is diagonal. Thus the order of the group element $(\mathfrak{Z}, S)$ is the order of $(\mathfrak{C}, S^t)$. To compute this latter order we need only show that when $\mathfrak{Z}$ is a field its degree over $\mathfrak{K}$ is the order of the group element $(\mathfrak{Z}, S)$. Let then $\mathfrak{Z}$ be a field and m the order of $(\mathfrak{Z}, S)$. Then since $(\mathfrak{W}_{-m}, T) = (\mathfrak{Z}, S)^m$ we have $\mathfrak{W}_{-m}$ diagonal. By Theorem 3 the expression of $\mathfrak{Z} \times \mathfrak{E}$ as the direct product of a cyclic field and a diagonal algebra is unique in the sense that if $\mathfrak{Z} \times \mathfrak{E} = \mathfrak{Z}_1 \times \mathfrak{E}_1$ then $\mathfrak{Z}_1$ is equivalent to $\mathfrak{Z}$ but $\mathfrak{E}_1 = \mathfrak{E}$. Since $\mathfrak{Z} \times \mathfrak{E} = \mathfrak{Z} \times \mathfrak{W}_{-tm}$ with $\mathfrak{W}_{-m}$ diagonal we have $\mathfrak{E} = \mathfrak{W}_{-m}$, $\mathfrak{E}$ is unaltered by ST^{-m}. Now $0 < m \leqq n$ and $\mathfrak{E}$ is unaltered by S and hence by T^{-m}. This is false unless $m = n$ since in particular $f_1^{T^{-m}} = f_{1-m} = f_{n+1-m} \neq f_1$ if $m < n$.

6. Similar cyclic systems

If $(\mathfrak{Z}, S)$ is a cyclic system we have $\mathfrak{Z} = \mathfrak{C} \times \mathfrak{E}$ where $\mathfrak{C}$ is a cyclic field with uniquely determined generating automorphism S^t. We then write $(\mathfrak{Z}, S) \sim (\mathfrak{C}, S^t)$ and say that $(\mathfrak{Z}, S)$ and $(\mathfrak{C}, S^t)$ are similar. If also $(\mathfrak{Z}_1, S_1) \sim (\mathfrak{C}_1, S_1^t) = (\mathfrak{C}, S^t)$ we call $(\mathfrak{Z}, S)$ and $(\mathfrak{Z}_1, S_1)$ similar and write $(\mathfrak{Z}_1, S_1) \sim (\mathfrak{Z}, S)$. This notation was also used by Teichmüller. We now prove

THEOREM 10. *Let $\mathfrak{Z}$ be a cyclic field of degree n over $\mathfrak{K}$, m be an integer, d be the g.c.d. of m and n so that $m = \mu d$, $n = \nu d$. Then if $\mathfrak{Y}$ is the cyclic subfield of degree ν over $\mathfrak{K}$ of $\mathfrak{Z}$ we have*

$$(\mathfrak{Z}, S)^m \sim (\mathfrak{Y}, S^\lambda), \tag{23}$$

where λ is an integer chosen so that $\lambda\mu \equiv 1 \pmod{\nu}$.

Observe that the generating automorphism S of $\mathfrak{Z}$ induces a generating automorphism S in $\mathfrak{Y}$, and we use the same notation in both. We first form $(\mathfrak{Z}, S)^d = (\mathfrak{W}_{-d}, T)$. The algebra $\mathfrak{W}_{-d} = (w_1, w_1^T, \cdots, w_1^{T^{n-1}})$ where, by (22),

$$w_1 = y_1 f_1 + y_2 f_{1-d} + \cdots + y_\nu f_{1-(\nu-1)d}, \tag{24}$$

and $y_1 = z_1 + z_{1+\nu} + \cdots + z_{1+(d-1)\nu}$. Clearly $y_1, \cdots, y_{\nu-1}$ form a normal basis of $\mathfrak{Y}$ over $\mathfrak{K}$; $y_i = y_1^{S^{i-1}}$. Every idempotent of $\mathfrak{Z} \times \mathfrak{E}$ is a sum of idempotents of $\mathfrak{E}$ and a simple computation shows that the only idempotents of $\mathfrak{E}$ unaltered by ST^{-d} are sums of the idempotents $g_1, g_1^T, \cdots, g_1^{T^{d-1}}$ where

$$g_1 = f_1 + f_{1-d} + \cdots + f_{1-(\nu-1)d}. \tag{25}$$

Hence $\mathfrak{W}_{-d} = \mathfrak{C} \times \mathfrak{E}_0$, $\mathfrak{E}_0 = (g_1, g_1^T, \cdots, g_1^{T^{d-1}})$ over $\mathfrak{K}$, $\mathfrak{C}$ is a cyclic field over $\mathfrak{K}$ with generating automorphism T^d. As in the proof of Theorem 3 we have

$$\mathfrak{W}_{-d} = \mathfrak{C}_1 \oplus \mathfrak{C}_1^T \oplus \cdots \oplus \mathfrak{C}_1^{T^{d-1}},$$

where $\mathfrak{C}_1 = \mathfrak{C}g_1$ and T^d induces a generating automorphism S_0 in $\mathfrak{C}_1$. Thus $(\mathfrak{C}, T^d) = (\mathfrak{C}_1, S_0)$. But w_1 is in $\mathfrak{C}_1$ and $(f_{1-d})^{T^d} = f_1$,

$$w_1^{S_0} = y_2 f_1 + \cdots + y_1 f_{1-(\nu-1)d} = w_1^S. \tag{26}$$

It follows that $\mathfrak{C}_1 = (w_1, w_1^{S_0}, \cdots, w_1^{S_0^{\nu-1}})$ over $\mathfrak{K}$, $(\mathfrak{C}, S_0) = (\mathfrak{Y}, S) = (\mathfrak{C}, T^d)$. Hence $(\mathfrak{W}_{-d}, T) \sim (\mathfrak{Y}, S)$, $(\mathfrak{Z}, S)^d \sim (\mathfrak{Y}, S)$. Also $(\mathfrak{Z}, S)^m = (\mathfrak{Z}, S)^{d\mu} \sim (\mathfrak{Y}, S)^\mu$ where μ is prime to the degree ν of $\mathfrak{Y}$. We have reduced the proof of our theorem to the case where m is prime to n, and wish to prove that $(\mathfrak{Z}, S)^m = (\mathfrak{Z}, S^q)$ where $qm \equiv 1 \pmod{n}$.

We form $(\mathfrak{Z}, S)^m$ by forming $\mathfrak{Z} \times \mathfrak{E}$, and see that $(\mathfrak{Z}, S)^m = (\mathfrak{W}_{-m}, T)$. Since m is prime to n the integers $1, 1 - m, 1 - 2m, \cdots, 1 - (n-1)m$ are distinct. But these are the subscripts on the idempotents f_i of (21), (22), and the correspondence $w_i \leftrightarrow z_i$ induces a simple isomorphism of $\mathfrak{Z}$ and $\mathfrak{W}_{-m}$. Now if $qm \equiv 1 \pmod{n}$ we have $2 - qm \equiv 1 \pmod{n}$, $f_{2-qm} = f_1$. Hence $w_1^T = z_1 f_2 + z_2 f_{2-m} + \cdots + z_n f_{2-m(n-1)} = z_{q+1} f_1 + z_{q+2} f_{1-m} + \cdots + z_{q+n} f_{2-m(n-1)} \leftrightarrow z_{q+1} = z_1^{S^q}$. It follows that $(\mathfrak{W}_{-m}, T) = (\mathfrak{Z}, S)^m = (\mathfrak{Z}, S^q)$ as desired.

The result above is the one referred to in our introduction.

7. The case $n = p^e$

Cyclic fields of prime power order are usually of particular interest. For semi-fields we have

THEOREM 11. *A semi-field $\mathfrak{Z}$ of order $n = p^e$, p a prime, with an automorphism S of order n is a cyclic semi-field with generating automorphism S.*

For otherwise $\mathfrak{Z}$ would be S-reducible and have S-irreducible components $\mathfrak{B}_i = \mathfrak{C}_i \times \mathfrak{E}_i$ each of order $n_i < n$. The order of an automorphism of a field $\mathfrak{C}_i$ is at most its degree s_i and the order of S in $\mathfrak{B}_i$ is thus at most n_i. But S has order p^e so that the order of S in $\mathfrak{B}_i$ is $p^{e_i} < n$. It follows that the order of S is the maximum $p^{e_i} < p^e$, a contradiction. Hence $\mathfrak{Z}$ is a cyclic semi-field with generating automorphism S.

It is known[13] that every cyclic field of degree (order) $n = p_1^{e_1} \cdots p_t^{e_t}$ over $\mathfrak{K}$ with the p_i distinct primes is uniquely expressible as a direct product of cyclic

[13] Cf. *Modern Higher Algebra*, p. 186.

fields of degrees $p_i^{e_i}$ and conversely. By applying this result together with Corollary II we have immediately

THEOREM 12. *Every cyclic semi-field of order* $n = p_1^{e_1} \cdots p_t^{e_t}$ *over* $\mathfrak{K}$ *for distinct primes* p_i *is expressible uniquely (in the sense of equivalence) as a direct product of cyclic semi-fields of respective orders* $p_i^{e_i}$, *and conversely.*

The result above may be used to reduce problems on cyclic semi-fields to the case $n = p^e$. We shall not do this here but shall continue with the case of general order n.

8. Norms in $\mathfrak{Z}$

Let z be in a cyclic semi-field $\mathfrak{Z}$ and define the norm of z to be the product

$$N_{\mathfrak{Z}|\mathfrak{K}}(z) = zz^S \cdots z^{S^{n-1}}. \tag{27}$$

The quantity $z = c_1e_1 + \cdots + c_te_t$ with the e_i pairwise orthogonal idempotents such that $e_i = e_{i-1}^S$, $e_i^{S^t} = e_i$, and hence

$$zz^{S^1} \cdots z^{S^{t(s-1)}} = e_1N_{\mathfrak{C}|\mathfrak{K}}(c_1) + e_2N_{\mathfrak{C}|\mathfrak{K}}(c_2) + \cdots + c_tN_{\mathfrak{C}|\mathfrak{K}}(c_t) = N_0.$$

But then

$$N_{\mathfrak{Z}|\mathfrak{K}}(z) = N_0N_0^S \cdots N_0^{S^{t-1}} = N_{\mathfrak{C}|\mathfrak{K}}(c_1c_2 \cdots c_t) \quad \text{in} \quad \mathfrak{K}. \tag{28}$$

Hence if $\gamma = N_{\mathfrak{Z}|\mathfrak{K}}(z)$ we have $\gamma = N_{\mathfrak{C}|\mathfrak{K}}(c)$ for c in $\mathfrak{C}$. Conversely if $\gamma = N_{\mathfrak{C}|\mathfrak{K}}(c)$ we take $c_1 = c$, $c_2 = \cdots = c_t = 1$ and have (28). We have proved

THEOREM 13. *A quantity* γ *of* $\mathfrak{K}$ *is the norm of a quantity* z *of* $\mathfrak{Z} = \mathfrak{C} \times \mathfrak{E}$ *if and only if* γ *is the norm of a quantity of the cyclic field* $\mathfrak{C}$.

9. Generalized cyclic algebras

Teichmüller connected the theory of cyclic systems with that of cyclic algebras. To do this we first let $\mathfrak{Z}$ be a cyclic field of degree n over $\mathfrak{K}$ with generating automorphism S. Then we may define a cyclic algebra

$$\mathfrak{Q} = (\mathfrak{Z}, S, \gamma) \tag{29}$$

consisting of all quantities of the form

$$z_0 + z_1y + \cdots + z_{n-1}y^{n-1} \qquad (z_i \text{ in } \mathfrak{Z}), \tag{30}$$

such that

$$y^iz = z^{S^i}y^i, \qquad y^n = \gamma \neq 0 \text{ in } \mathfrak{K} \qquad (z \text{ in } \mathfrak{Z},\ i = 0, \cdots, n-1). \tag{31}$$

Such algebras are[14] always normal simple over $\mathfrak{K}$ and are total matric algebras if $\gamma = 1$.

We may now generalize such algebras by allowing $\mathfrak{Z}$ to be any cyclic semi-field. The resulting algebras arise in a very simple fashion. For we may prove the explicit result

[14] Ibid, p. 241.

THEOREM 14. *Let $\mathfrak{Z}$ be a cyclic semi-field with generating automorphism S so that $\mathfrak{Z} = \mathfrak{C} \times \mathfrak{E}$ where $\mathfrak{E}$ is a diagonal algebra of order t over $\mathfrak{K}$ with generating automorphism S_1 induced by S, $\mathfrak{C}$ is a cyclic field with generating automorphism S_2 induced by S^t. Then $\mathfrak{Q} = (\mathfrak{Z}, S, \gamma)$ if and only if*

$$\mathfrak{Q} = (\mathfrak{C}, S_2, \gamma) \times (\mathfrak{E}, S_1, \gamma). \tag{32}$$

For it is known[15] that $(\mathfrak{E}, S_1, \gamma)$ is a total matric algebra for any $\gamma \neq 0$. Moreover $\mathfrak{M} = (\mathfrak{E}, S_1, \gamma)$ has an ordinary matric basis e_{ij} with $e_{ii} = e_i$ where $e_i = e_{i-1}^{S_1}$, $\mathfrak{E} = (e_1, \cdots, e_t)$. We let y_0 be the quantity y of (30), (31) for the algebra $(\mathfrak{C}, S_2, \gamma)$ and set

$$y = e_{21} + e_{32} + \cdots + e_{t,t-1} + y_0 e_{1t}. \tag{33}$$

Then y is in the algebra $\mathfrak{M} \times \mathfrak{K}[y_0]$. The general quantity of $\mathfrak{Z} = \mathfrak{C} \times \mathfrak{E}$ has the form $z = c_1 e_1 + \cdots + c_t e_t$ for c_i in $\mathfrak{C}$ and $z^S = c_1 e_2 + \cdots + c_t^{S_2} e_1$. But $ye_i = e_{i+1}y$, $jc = c^{S_2}j$, so that $yz = z^S y$, $y^i z = z^{S^i} y^i$, $y^t = j$, $y^n = \gamma$. A trivial computation shows that the set of all quantities (30) has order n^2 over $\mathfrak{K}$, $(\mathfrak{C}, S_2, \gamma) \times (\mathfrak{E}, S_1, \gamma) = (\mathfrak{Z}, S, \gamma)$. Conversely if $\mathfrak{Q} = (\mathfrak{Z}, S, \gamma)$ we form $(\mathfrak{C}, S_2, \gamma) \times (\mathfrak{E}, S_1, \gamma)$ and obtain an algebra equivalent to $\mathfrak{Q}$.

Thus every generalized cyclic algebra is a normal simple algebra similar to an ordinary cyclic algebra. If $\mathfrak{Q}_1$ and $\mathfrak{Q}_2$ are two generalized cyclic algebras the meaning of $\mathfrak{Q}_1 \sim \mathfrak{Q}_2$ is the usual one for normal simple algebras. We now have the following extensions of corresponding results on ordinary cyclic algebras.

THEOREM 15. *The algebra $(\mathfrak{Z}, S, 1)$ is a total matric algebra.*

For $(\mathfrak{Z}, S, 1) \sim (\mathfrak{C}, S^t, 1)$ where $\mathfrak{C}$ is a cyclic field and as we have said $(\mathfrak{C}, S^t, 1)$ is a total matric algebra.

THEOREM 16. *The algebra $\mathfrak{Z}$ is a maximal commutative subalgebra of $(\mathfrak{Z}, S, \gamma)$.*

For let $az = za$ for every z of $\mathfrak{Z}$. We write $(\mathfrak{Z}, S, \gamma) = \mathfrak{M} \times (\mathfrak{C}, S^t, \gamma)$ where $\mathfrak{M}$ is a total matric algebra with a basis $e_{ij} (i, j = 1, \cdots, t)$. Then $a = \sum a_{ij} e_{ij}$ with e_{ij} in $(\mathfrak{C}, S^t, \gamma)$. Now $ac = ca$ for every c of $\mathfrak{C}$, $\sum (a_{ij}c - ca_{ij})e_{ij} = 0$. It follows that the a_{ij} are each commutative with every c of $\mathfrak{C}$. Since $\mathfrak{C}$ is known[16] to be maximal in $(\mathfrak{C}, S^t, \gamma)$ we have the a_{ij} in $\mathfrak{C}$. Thus a is in the algebra $\mathfrak{M}$ over $\mathfrak{C}$ and is commutative with all of its diagonal matrices, the elements of $\mathfrak{E}$, a must be diagonal, a is in $\mathfrak{Z}$.

By the use of Theorems 15, and 16 we may prove

THEOREM 17. *The algebra $(\mathfrak{Z}, S, \gamma)$ is a total matric algebra if and only if $\gamma = N_{\mathfrak{Z}|\mathfrak{K}}(z)$ for z in $\mathfrak{Z}$.*

However the result also follows from Theorems 13, 14. For $(\mathfrak{Z}, S, \gamma) \sim (\mathfrak{C}, S^t, \gamma)$ is a total matric algebra if and only if $\gamma = N_{\mathfrak{C}|\mathfrak{K}}(c)$, c in $\mathfrak{C}$. But this result is equivalent to that of Theorem 17 by Theorem 13.

By replacing y of (30) by y^λ for λ prime to n we replace γ by γ^λ, S by S^λ and retain all our properties of $\mathfrak{Q}$. This gives

---

[15] Loc. cit., pp. 91, 92.

[16] Loc. cit., p. 243.

THEOREM 18. *Let* λ *be prime to* n. *Then* $(\mathfrak{Z}, S, \gamma) = (\mathfrak{Z}, S^\lambda, \gamma^\lambda)$.

Theorem 16 also has the immediate consequence

THEOREM 19. *The algebra* $(\mathfrak{Z}, S, \gamma) = (\mathfrak{Z}, S, \delta)$ *if and only if* $\delta = \gamma N_{\mathfrak{Z}|\mathfrak{F}}(z)$ *for* z *in* $\mathfrak{Z}$.

COROLLARY. $(\mathfrak{Z}, S, \gamma) = (\mathfrak{Z}, S, \delta)$ *if and only if* $\delta = \gamma N_{\mathfrak{C}|\mathfrak{F}}(c)$ *for* c *in the maximal cyclic subfield* $\mathfrak{C}$ *of* $\mathfrak{Z}$.

Cyclic algebras are not defined in terms of cyclic fields but rather in terms of given cyclic systems. That this is true is evidenced by Theorem 18 and also by the notation $(\mathfrak{Z}, S, \gamma)$. We shall simplify the writing of some of our theorems by introducing the notation

$$\mathfrak{Q} = (\mathfrak{A}, \gamma) \qquad (\gamma \neq 0 \text{ in } \mathfrak{F}), \tag{34}$$

for the algebra $\mathfrak{Q} = (\mathfrak{Z}, S, \gamma)$, $\mathfrak{A} = (\mathfrak{Z}, S)$. Thus if $\mathfrak{C}$ is a diagonal algebra of order n so that $\mathfrak{J} = (\mathfrak{C}, S)$ is the identity of the group of cyclic systems of degree n the algebra $(\mathfrak{J}, \gamma)$ is a total matric algebra for every $\gamma \neq 0$ in $\mathfrak{F}$.

10. Applications of theorems on cyclic systems

The result given by Theorem 14 was obtained by Teichmüller and expressed in the form

THEOREM 20. *Let* $\mathfrak{A} \sim \mathfrak{B}$. *Then* $(\mathfrak{A}, \gamma) \sim (\mathfrak{B}, \gamma)$.

Teichmüller also proved the important

THEOREM 21. *Let* $\mathfrak{A}$ *and* $\mathfrak{B}$ *be cyclic systems of the same degree. Then*

$$(\mathfrak{A}, \gamma) \times (\mathfrak{B}, \gamma) \sim (\mathfrak{A} \cdot \mathfrak{B}, \gamma). \tag{35}$$

We may prove this result as a consequence of the more general[17]

THEOREM 22. $(\mathfrak{A}, \gamma) \times (\mathfrak{B}, \delta) = (\mathfrak{A}, \gamma\delta^{-1}) \times (\mathfrak{A} \cdot \mathfrak{B}, \delta)$.

For if we put $\gamma = \delta$ in Theorem 22 we obtain the total matric algebra $(\mathfrak{A}, 1)$ and have Theorem 21. To prove Theorem 22 let $\mathfrak{A} = (\mathfrak{Z}, S)$ and as in (30), (31), define y in $(\mathfrak{A}, \gamma)$ such that $yz = z^S y$ for every z of $\mathfrak{Z}$. Similarly let $\mathfrak{B} = (\mathfrak{B}, T)$ and j be in $(\mathfrak{B}, \delta)$ such that $jv = v^T j$ for every v of $\mathfrak{B}$. Then $\mathfrak{Z} \times \mathfrak{B} = \mathfrak{Z} \times \mathfrak{W}_{-1}$ where $\mathfrak{A} \cdot \mathfrak{B} = (\mathfrak{W}_{-1}, T)$. Now $\mathfrak{Q} = (\mathfrak{A}, \gamma) \times (\mathfrak{B}, \delta)$ contains the algebra $\mathfrak{Q}_1 = (\mathfrak{Z}, ST^{-1}, \gamma\delta^{-1})$ generated by $\mathfrak{Z}$ and yj^{-1} and also the algebra $\mathfrak{Q}_2 = (\mathfrak{W}_{-1}, T, \delta) = (\mathfrak{A} \cdot \mathfrak{B}, \delta)$ generated by $\mathfrak{W}_{-1}$ and j. Clearly every quantity of $\mathfrak{Q}_1$ is commutative with every quantity of $\mathfrak{Q}_2$ and a simple computation shows that the only quantities of $\mathfrak{Q}$ commutative with every quantity of $\mathfrak{Q}_1$ are those of $\mathfrak{Q}_2$. Since $\mathfrak{Q}_1$ is normal simple, $\mathfrak{Q} = \mathfrak{Q}_1 \times \mathfrak{Q}_2$ as desired.

The following result is an immediate corollary of Theorem 21.

THEOREM 23. $(\mathfrak{A}, \gamma)^m \sim (\mathfrak{A}^m, \gamma)$.

In considering the direct powers of $(\mathfrak{A}, \gamma)$ we may use Theorem 20 to reduce the study to the case where $\mathfrak{A} = (\mathfrak{Z}, S)$ with $\mathfrak{Z}$ a field. Hence $\mathfrak{A}$ has degree and order n and we apply Theorem 10 to Theorem 18 and obtain

---

[17] For an analogous result for the case where n is a power of the characteristic p of $\mathfrak{F}$ see E. Witt, Journal für Mathematik, vol. 176 (1936), pp. 126-140, p. 138.

THEOREM 24. *Let $\mathfrak{A}$ have degree and order n and let μ be prime to n, $\lambda\mu \equiv 1 \pmod{n}$. Then $(\mathfrak{A}, \gamma) = (\mathfrak{A}^{\mu}, \gamma^{\lambda})$.*

We next apply Theorem 10 to Theorem 23 to obtain the principal result

THEOREM 25. *Let d be the g.c.d. of n and $m = \mu d$, $\mathfrak{A} = (\mathfrak{Z}, S)$ be a cyclic system of degree and order $n = d\nu$ over $\mathfrak{K}$, $\mathfrak{Y}$ be the cyclic subfield of $\mathfrak{Z}$ of degree ν over $\mathfrak{K}$. Then*

$$(\mathfrak{A}, \gamma)^{m} \sim (\mathfrak{B}, \gamma) \tag{36}$$

where $\mathfrak{B} = (\mathfrak{Y}, S^{\lambda})$, $\lambda\mu \equiv 1 \pmod{\nu}$.

We may also prove

THEOREM 26. $(\mathfrak{A}, \gamma) \times (\mathfrak{A}, \delta) \sim (\mathfrak{A}, \gamma\delta)$.

For by Theorem 20 it is sufficient to consider the case $\mathfrak{A} = (\mathfrak{Z}, S)$ with $\mathfrak{Z}$ a field. Take $\mu = n - 1$ in Theorem 24 and have $\lambda = -1$, $(\mathfrak{A}, \delta) = (\mathfrak{A}^{n-1}, \delta^{-1})$. Then $(\mathfrak{A}, \gamma) \times (\mathfrak{A}, \delta) = (\mathfrak{A}, \gamma) \times (\mathfrak{A}^{n-1}, \delta^{-1})$. Apply Theorem 22 to obtain $(\mathfrak{A}, \gamma) \times (\mathfrak{A}, \delta) = (\mathfrak{A}, \gamma\delta) \times (\mathfrak{A}^{n}, \delta^{-1}) \sim (\mathfrak{A}, \gamma\delta)$ since $\mathfrak{A}^{n} = \mathfrak{Z}$, $(\mathfrak{Z}, \gamma)$ is a total matric algebra.

As a final consequence we have

THEOREM 27. *Let $\mathfrak{Y}$ be as in Theorem 25. Then*

$$(\mathfrak{Z}, S, \gamma) \times (\mathfrak{Y}, S, \delta) \sim (\mathfrak{Z}, S, \gamma\delta^{d}).$$

For $(\mathfrak{Y}, S, \delta) \sim (\mathfrak{Z}, S, \delta)^{d} \sim (\mathfrak{Z}, S, \delta^{d})$ by Theorems 25 and 26. It follows that $(\mathfrak{Z}, S, \gamma) \times (\mathfrak{Y}, S, \delta) \sim (\mathfrak{Z}, S, \gamma) \times (\mathfrak{Z}, S, \delta^{d}) \sim (\mathfrak{Z}, S, \gamma\delta^{d})$ by Theorem 26.

In closing observe that we have proved certain theorems on cyclic algebras as a consequence of Theorems 20 and 21 and our theorems on cyclic systems. Teichmüller worked in the converse direction and only succeeded in proving his results by the use of the following rather ingenious device. Let ξ be an indeterminate over $\mathfrak{K}$, $\Omega = \mathfrak{K}(\xi)$, $\mathfrak{A}_{\Omega}$ designate the cyclic system $(\mathfrak{Z}_{\Omega}, S)$. Then $(\mathfrak{A}_{\Omega}, \xi)$ is a total matric algebra if and only if $\mathfrak{A}$ has order unity, that is, $\mathfrak{Z}$ is a diagonal algebra. It follows that $(\xi, \mathfrak{A}_{\Omega}) \times (\xi, \mathfrak{B}_{\Omega}) \sim (\xi, \mathfrak{C}_{\Omega})$ if and only if $\mathfrak{C} = \mathfrak{A} \cdot \mathfrak{B}$. Teichmüller was then able to base his proofs of the properties of cyclic systems upon the equivalent ones of cyclic algebras.

THE UNIVERSITY OF CHICAGO.

ANNALS OF MATHEMATICS
Vol. 39, No. 3, July, 1938

ON A GENERALIZATION OF KALUZA'S THEORY OF ELECTRICITY

BY A. EINSTEIN AND P. BERGMANN

(Received April 8, 1938)

INTRODUCTION

So far, two fairly simple and natural attempts to connect gravitation and electricity by a unitary field theory have been made, one by Weyl, the other by Kaluza. Furthermore, there have been some attempts to represent Kaluza's theory formally so as to avoid the introduction of the fifth dimension of the physical continuum. The theory presented here differs from Kaluza's in one essential point; we ascribe physical reality to the fifth dimension whereas in Kaluza's theory this fifth dimension was introduced only in order to obtain new components of the metric tensor representing the electromagnetic field. Kaluza assumes the dependence of the field variables on the four coördinates x^1, x^2, x^3, x^4 and not on the fifth coördinate x^0 when a suitable coördinate system is chosen.

It is clear that this is due to the fact that the physical continuum is, according to our experience a four dimensional one. We shall show, however, that it is possible to assign some meaning to the fifth coördinate without contradicting the four dimensional character of the physical continuum.

In the first chapter of our paper we present Kaluza's original theory; in the second, its new generalization. This is done in order to make the reading easier.

In the appendix we simplify the derivation of the field equations by generalizing the tensor calculus for the case of tensor densities.

I. THE KALUZA THEORY[1]

1. We consider a five dimensional space (4 + 1 dimensions) with the metric

$$d\sigma^2 = \gamma_{\mu\nu} dx^\mu dx^\nu. \tag{1}$$

2. We assume that the space is "cylindrical" with respect to a definite vector field. Analytically this means: There exists a contravariant vector A^ν so that if τ is an infinitesimally small constant, then

$$\delta x^\nu = \tau \cdot A^\nu \tag{2}$$

is an infinitesimal displacement of the space points. The element must not

[1] Contained in part in O. Klein, Zeitschrift f. Phys., **37** (1926), 895-906, A. Einstein, Sitzungsberichte d. Preuss. Akademie d. Wissenschaften, Physik.-mathemat. Klasse, 1927, 23-30.

change its length given by (1) if its endpoints are displaced according to (2). This is expressed by the equation:

$$\gamma_{\mu\nu,\sigma}A^{\sigma} + \gamma_{\mu\sigma}A^{\sigma}_{,\nu} + \gamma_{\nu\sigma}A^{\sigma}_{,\mu} = 0. \tag{3a}$$

The invariant character of this equation can easily be shown. Putting

$$\gamma_{\mu\sigma}A^{\sigma} = A_{\mu}, \tag{3b}$$

we get

$$A_{\mu,\nu} + A_{\nu,\mu} - A^{\sigma}\{\gamma_{\mu\sigma,\nu} + \gamma_{\nu\sigma,\mu} - \gamma_{\mu\nu,\sigma}\} = 0, \tag{3c}$$

or

$$\{A_{\mu,\nu} - A_{\sigma}\Gamma^{\sigma}_{\mu\nu}\} + \{A_{\nu,\mu} - A_{\sigma}\Gamma^{\sigma}_{\nu\mu}\} = 0, \tag{3d}$$

where $\Gamma^{\lambda}_{\iota\kappa}$ are the Christoffel symbols of the second kind. Or, in the language of the absolute tensor calculus we have:

$$A_{\mu;\nu} + A_{\nu;\mu} = 0.^{2} \tag{3}$$

It follows from (3) that $\gamma_{\mu\nu}A^{\mu}A^{\nu}$ (that is the norm of A^{μ}) is constant along the lines to which A^{μ} is a tangent. By multiplying (3) by $A^{\mu}A^{\nu}$ we obtain:

$$A^{\nu}A^{\mu}A_{\mu;\nu} + A^{\mu}A^{\nu}A_{\nu;\mu} = A^{\nu}(A^{\mu}A_{\mu})_{,\nu} = 0. \tag{4}$$

3. Kaluza's theory imposes another property on the vector field A^{μ}, besides the one expressed in (3): The lines to which the A^{μ} are tangents—the "A-lines"—have to be geodesics. Analytically this means: The lines defined by

$$\frac{dx^{\nu}}{d\sigma} = \lambda A^{\nu} \tag{5}$$

(where $1/\lambda^2$ is equal to the norm of A) satisfy the equation

$$\frac{d^2x^{\nu}}{d\sigma^2} + \Gamma^{\nu}_{\alpha\beta}\frac{dx^{\alpha}}{d\sigma}\frac{dx^{\beta}}{d\sigma} = 0, \tag{6a}$$

or, according to (5):

$$\frac{\partial \ln \lambda}{\partial x^{\alpha}}A^{\alpha}A^{\nu} + A^{\alpha}A^{\nu}_{;\alpha} = 0. \tag{6b}$$

The first term vanishes because of (4) $\left(\frac{\partial \ln \lambda}{\partial x^{\alpha}} \cdot A^{\alpha}\right.$ is the derivative of a function of the norm of A in the direction of $\left.A\right)$. Therefore:

$$A^{\alpha}A^{\nu}_{;\alpha} = 0, \tag{6c}$$

---

[2] Killing's equation.

or

$$A^{\alpha}A_{\nu;\alpha} = 0. \tag{6d}$$

From (3) follows:

$$A^{\alpha}A_{\alpha;\nu} = \tfrac{1}{2}(A^{\alpha}A_{\alpha})_{,\nu} = 0. \tag{6}$$

This means that the norm of A is constant not only along the A-lines but in the whole space. We characterize, in this way, the space structure in Kaluza's theory.

REMARK: From the "cylindrical" character of the space there follows the existence of a vector A the *symmetrical* derivatives of which vanish. But there remain the *antisymmetrical* derivatives of A_{μ} which represent, in Kaluza's theory, the electromagnetic field.

The Special Coördinate System

We consider a four dimensional surface cutting each of the A-lines once and only once. We introduce on this surface 4 coördinates x^{a} $(a = 1 \cdots 4)$ and assume x^{0} equal zero on this surface. Through each point of the surface passes an A-line on which one of the two directions is chosen as positive; this choice is to be the same on all A-lines. As x^{0}-coördinate we choose the distance along an A-line calculated with the help of (1), starting from the original four dimensional surface with $x^{0} = 0$, the chosen direction determining the sign of x^{0}.

Therefore, on the x^{0}-line:

$$x^{0} = \int_{0}^{x^{0}} \sqrt{\gamma_{00}}\, dx^{0}, \tag{7a}$$

and therefore in the whole space,

$$\gamma_{00} = 1. \tag{7}$$

As the absolute value of A is always 1, we have

$$\gamma_{00} A^{0}A^{0} = 1, \tag{8a}$$

or,

$$A^{0} = 1, \tag{8}$$

since, because of the choice of our coördinate system, $A^{1} = A^{2} = A^{3} = A^{4} = 0$.

From this and (3a), expressing the condition of cylindricity, we have

$$\gamma_{\mu\nu,0} = 0. \tag{9}$$

As generally

$$A_{\mu} = \gamma_{\mu\nu}A^{\nu},$$

we have in the special coördinate system,

$$A_m = \gamma_{0m}, \qquad A_0 = \gamma_{00} = 1. \tag{10}$$

None of the components of A_μ depend on x^0 because of (9). Therefore in the antisymmetrical derivatives

$$A_{\mu\nu} = A_{\mu,\nu} - A_{\nu,\mu}$$

all components bearing one index zero vanish identically.

In our special coördinate system the space structure is described by 10 functions γ_{mn} and 4 functions $\gamma_{0m}(= A_m)$, which do not depend on x^0. (Here and later the latin indices run only from 1 to 4.) This is, therefore, a four dimensional description of the space. But the variables introduced are not, from the point of view of our description, the most natural ones, as will be shown soon.

We ask: What coördinate transformations are still allowed preserving our special coördinate system?

a. The coördinates x^a have been chosen arbitrarily on the four dimensional surface (with $x^0 = 0$). Therefore the following transformation is still allowed:

$$\begin{aligned} \bar{x}^a &= \bar{x}^a(x^1 \cdots x^4), \\ \bar{x}^0 &= x^0. \end{aligned} \tag{11}$$

We call it the *Four-transformation.*

b. The four dimensional surface (with $x^0 = 0$) was chosen arbitrarily. It can be replaced by another surface without changing the x^a-coördinates corresponding to the A-lines. The appropriate transformation is:

$$\begin{aligned} \bar{x}^a &= x^a, \\ \bar{x}^0 &= x^0 + f(x^1 \cdots x^4). \end{aligned} \tag{12}$$

We call it the *Cut-transformation.*

The γ_{mn} behave, with respect to a four-transformation, as a four dimensional tensor and $A_n(= \gamma_{0n})$ as a four-vector. With respect to the cut-transformation we have

$$\bar{A}_m = A_m - \frac{\partial f}{\partial x^m}, \tag{13}$$

$$\bar{\gamma}_{mn} = \gamma_{mn} - \gamma_{m0}\frac{\partial f}{\partial x^n} - \gamma_{n0}\frac{\partial f}{\partial x^m} + \frac{\partial f}{\partial x^m}\frac{\partial f}{\partial x^n}. \tag{14}$$

Instead of the γ_{mn} which are not invariant with respect to cut-transformations (according to (14)) we introduce new functions g_{mn} by the following consideration: Let two infinitesimally close A-lines L and L' pass through two infinitesimally close points P and P'. There exists between L and L' an

extremal distance because of the existence of a metric. This distance is determined by

$$g_{\mu\nu} = \gamma_{\mu\nu} - A_\mu A_\nu. \tag{15}$$

The $g_{\mu\nu}$ defined in this way are components of a five dimensional tensor. In the special coördinate system only those of its components which have no zero index do not vanish, i.e. the g_{mn} are its only components. The g_{mn} behave like a four dimensional tensor with respect to the four-transformation and are invariant with respect to cut-transformations.

It is, therefore, quite natural to choose g_{mn} and A_m as field variables in the special coördinate system; the g_{mn} are the components of the physical metrical tensor. The antisymmetrical derivatives of A_m, that is $A_{m,n} - A_{n,m}$, are invariant with respect to the cut-transformation, but not the components A_m themselves. This corresponds to the fact that the electromagnetic potentials are defined only up to additive terms which are gradients of an arbitrary function.

This result can be summarized: The five dimensional space structure is equivalent to a four dimensional one with a metric (g_{mn}) and a vector field (A_m), determined only up to an additive arbitrary gradient.

Kaluza's roundabout way of introducing the five dimensional continuum allows us to regard the gravitational and electromagnetic fields as a unitary space structure. This result is essential. At first sight it could be argued: Is it really a step forward to introduce a five dimensional metric and a vector for which we assume some arbitrary restrictions instead of a four dimensional metric and a four-vector? But to be just to Kaluza's theory it should be born in mind that the one really arbitrary step is taken only when the five dimensional representation of the four dimensional space is assumed. If, however, this is done, then the introduction of a five-vector is merely a consequence, necessary because of the four dimensional character of the empirical space. The representation of the electromagnetic field by a potential vector, achieved in this way, is certainly not a trivial result.

But apart from this point, the result is rather disappointing. Kaluza's aim was undoubtedly to obtain some new physical aspect for gravitation and electricity by introducing a unitary field structure. This end was, however, not achieved.

The field equations of the theory can be derived for instance with the help of a variational principle. This means a choice of an action function $\mathfrak{H}$, which must be a scalar density, i.e. $\frac{1}{\sqrt{g}} \cdot \mathfrak{H}$ must be an invariant with respect to the four- and cut-transformations. Therefore, in our special coördinate system, the invariant has to be formed of the g_{mn} and A_m as well as of such derivatives of these quantities which do not change if A_m is replaced by $A_m - \partial f/\partial x^m$. We can conclude, therefore, that A_m can enter the action function only through their antisymmetric derivatives. If the Hamiltonian $\mathfrak{H}$ has to contain only

second derivatives of the potentials g_{mn}, A_m, or expressions of the second order containing only first derivatives, then the most general action function is of the form

$$\mathfrak{H} = \sqrt{g}(R + c \cdot M), \tag{16}$$

where R is the curvature invariant and M the Maxwell invariant.

Many fruitless efforts to find a field representation of matter free from singularities based on this theory have convinced us, however that such a solution does not exist.[3]

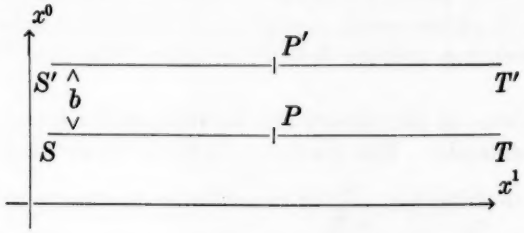
II. Generalization of Kaluza's Theory

General Remarks about the Space Structure

If Kaluza's attempt is a real step forward, then it is because of the introduction of the five dimensional space. There have been many attempts to retain the essential formal results obtained by Kaluza without sacrificing the four dimensional character of the physical space. This shows distinctly how vividly our physical intuition resists the introduction of the fifth dimension. But by considering and comparing all these attempts one must come to the conclusion that all these endeavors did not improve the situation. It seems impossible to formulate Kaluza's idea in a simple way without introducing the fifth dimension.

We have, therefore, to take the fifth dimension seriously although we are not encouraged to do so by plain experience. If, therefore, the space structure seems to force acceptance of the five dimensional space theory upon us we must ask whether it is sensible to assume the rigorous reducibility to four dimensional space. We believe that the answer should be "no," provided that it is possible to understand, in another way, the quasi-four dimensional character of the physical space by taking as a basis the five dimensional continuum and to simplify hereby the basic geometrical assumptions.

To make the underlying ideas clearer we consider a two dimensional space (instead of the five dimensional one), that is approximately a one dimensional continuum (instead of a four dimensional one).



[3] We tried to find a rigorous solution of the gravitational equations, free from singularities, by taking into account the electromagnetic field. We thought that a solution of a rotational-symmetrical character could, perhaps, represent an elementary particle. Our investigation was based on the theory of "bridges" (Einstein and Rosen, Phys. Rev., 48: 73 (1935). We convinced ourselves, however, that no solution of this character exists.

The space considered is the narrow strip limited by the lines ST and $S'T'$, The coördinates are x^0 and x^1, and, for the sake of simplicity, the Euclidean character of the space is assumed. We imagine the strip curved into a tube so that ST coincides with $S'T'$. Every point P on ST coincides in this way with a certain point P' on $S'T'$, so that a small cylindrical surface is formed. If the width of the strip, that is, the circumference of the cylinder (denoted by b), is small, and if a continuous and slowly changing function $\varphi(x^0, x^1)$ is given, that is if $b \cdot \partial\varphi/\partial x^a$ is small compared with φ, then the values of φ belonging to the points on the segment PP' differ from each other very slightly and φ can be regarded, approximately, as a function of x^1 only. This reduction of the number of dimensions of the space was achieved because the space is closed in the x^0-dimension and the "width" is very small. But this quasi-one dimensional character of the space does not exist if the function φ varies too rapidly.

The assumption concerning the Euclidean character is certainly unessential for our consideration. It may be essential for our argument that some metric is assumed so that it is possible to speak of a circumference in the x^0-direction, but in any case we are interested only in spaces with a Riemannian metric.

It is more convenient to consider, in the following, instead of a space "closed" in the x^0-direction[4] a space "periodic" in the x^0-direction. The "periodic" and the "closed" character are equivalent if the corresponding points $P, P', P'', \cdots$ are regarded as "the same" point.

By putting in our picture the four dimensional continuum $x^1 \cdots x^4$ instead of the one dimensional continuum x^1 we obtain a model of the physical space which will form the basis of our future considerations.

The most essential point of our theory is the replacing of the hypothesis 2, of rigorous cylindricity by the assumption that space is closed (or periodic) in the x^0-direction.

The Space Structure

We shall characterize the space according to our previous remarks concerning the modified form of Kaluza's theory.

1. We consider a five dimensional space with the metric (1).

2. The space is closed with respect to one dimension. This closed space will be represented by a space which is open and periodic with respect to this dimension. A point P of the physical space will, therefore, be represented by an infinite number of points $P, P', P'', \cdots$.

REMARK: This postulate replaces the cylindricity condition in Kaluza's original theory.

3. Through every space point there passes a geodesic line closed in itself and free from singularities. Or, formulated for the "periodic" space: If two points P, P' corresponding to each other, but otherwise arbitrary, are connected by

[4] This expression is not quite clear but it may be used here in a chapter aiming only at a general orientation.

the geodesic line,[5] then this line also passes through all other points corresponding to P, i.e. through P'', P''' etc.

This axiom is analogous to the axiom 3. of the original Kaluza theory. The geodesic lines introduced here correspond to the A-lines in Kaluza's theory and we shall again call them the A-lines.

We again introduce the special coördinate system. We consider a four dimensional surface $x^0 = 0$ cutting all A-lines once and only once. The four coördinates $x^1 \cdots x^4$ on this surface determine a point on this surface as well as an A-line through this point. Any point P on this A-line is determined by the distance (measured along this line with the help of (1)) between $\overset{0}{P}$ (that is the intersecting point of the A-line and the surface $x^0 = 0$) and P. For points on the "positive" side of the surface $x^0 = 0$ we have

$$x^0 = \frac{1}{b}\int_{\overset{0}{P}}^{P} d\sigma, \tag{17}$$

where b is

$$b = \int^{P'} d\sigma. \tag{17a}$$

Here P, P' are two corresponding, consecutive points lying on the A-line. Therefore, b depends on $x^1 \cdots x^4$ only. By this convention the difference of the x^0-coördinates of two corresponding, consecutive points (which, of course, lie on the same x^0-line) equals 1.

By the assumptions thus made five coördinates x^0, $x^1 \cdots x^4$ correspond to every space point.

Lines with constant $x^1 \cdots x^4$ are geodesic lines. On these lines we have:

$$\frac{dx^1}{d\sigma} = \frac{dx^2}{d\sigma} = \frac{dx^3}{d\sigma} = \frac{dx^4}{d\sigma} = 0, \qquad \frac{dx^0}{d\sigma} = \frac{1}{b}, \qquad \frac{d^2x^0}{d\sigma^2} = 0. \tag{18a}$$

Therefore, the equation of the geodesic line gives

$$\Gamma^{\alpha}_{00} = 0. \tag{18b}$$

From this follows:

$$\begin{bmatrix} \alpha \\ 00 \end{bmatrix} = 0, \tag{18c}$$

or

$$\gamma_{00,0} = 0 \tag{18d}$$

and

$$2\gamma_{\alpha 0,0} - \gamma_{00,\alpha} = 0. \tag{18}$$

[5] We do not consider the existence of more than one geodesic connection.

Equation (18d) does not contain anything new because we have on the A-lines:

(19a) $$d\sigma^2 = b^2 dx^{0^2}$$

and

(19) $$\gamma_{33} = b^2$$

which is independent of x^0.

We integrate (18) over one period along the A-line. The integral of the first term on the left hand side in (18) gives zero because of the periodic character of the space. Because of (17) and (19) we obtain:

(20) $$\frac{\partial}{\partial x^a}\int_{P'}^{P''} \gamma_{00} dx^0 = \frac{\partial}{\partial x^a}\left\{b\int_{P'}^{P''} d\sigma\right\} = \frac{\partial}{\partial x^a}(b^2) = 0.$$

Therefore b and γ_{00} are independent of $x^1 \cdots x^4$ too, i.e. constant in the whole space. We may write:

(21) $$\begin{cases} \gamma_{00} = 1, \\ b = 1. \end{cases}$$

In a strict sense we were not justified in choosing the value for γ_{00} as we have already chosen the coördinate difference Δx^0 of corresponding, consecutive points to be equal 1. We take this fact into account by putting Δx^0 not 1, but some small quantity λ. Then, because of (21), λ is simultaneously the metric distance $P'P''$.

Furthermore, from (18) follows because of (21)

(22) $$\frac{\partial}{\partial x^0}\gamma_{a0} = 0.$$

Therefore, in the generalized theory also, the γ_{0a} (but not the γ_{ab}) are independent of x^0. We can introduce, in the generalized theory as well as before, the field of contravariant unit vectors A^μ the components of which are, in the special coördinate system, 0, 0, 0, 0, 1. Therefore again

(22a) $$A_a = A^\beta \gamma_{\beta a} = \gamma_{0a}.$$

We can also introduce as the metric tensor instead of the

(23) $$g_{\alpha\beta} = \gamma_{\alpha\beta} - A_\alpha A_\beta,$$

of which only those components which have no index zero do not vanish. There exists, in this sense, here also a four dimensional metrical tensor g_{ab}. *Its components are, however, in general periodic functions of* x^0. The whole difference between the new theory and Kaluza's lies just in this fact. (A_m in both theories do not depend on x^0.)

Having characterized the space structure we shall turn later, after some mathematical preparations, to the formulation of field equations which can be

derived from a variational principle in the space considered. We shall again demand that the Hamiltonian be composed additively of terms containing either second derivatives (linearly) or products of two first derivatives.

Tensor Analysis with Respect to the Special Coördinate System

From our definition of the special coördinate system it follows that again there exists the freedom of "four-transformations" and "cut-transformations." In order to avoid the formal difficulties of general coördinate systems we have to investigate the transformation properties of the quantities formed with respect to the four- and cut-transformations. Definition of the contravariant four-vector: Four quantities a^σ (being functions of x^0, $x^1 \cdots x^4$) form a contravariant vector if, with respect to four-transformations, the following transformation law is valid:

$$\bar{a}^\sigma = \frac{\partial \bar{x}^\sigma}{\partial x^\iota} a^\iota, \tag{24}$$

and if the a^σ are invariant with respect to cut-transformations. The definition of a covariant vector is quite analogous. The definition of a tensor also follows from this.

Therefore g_{mn} is a tensor. However $A_m (= \gamma_{0m})$ is not a vector as the components transform, with respect to a cut-transformation, according to (13). The $A_{mn} = A_{m,n} - A_{n,m}$ are, however components of a tensor.

The tensor algebra is the same as the normal four dimensional one and does not need to be carried out here.

We now direct our attention to the formation of new tensors by differentiation.

1. For both four- and cut-transformations we have

$$\frac{\partial}{\partial \bar{x}^0} = \frac{\partial}{\partial x^0}. \tag{25}$$

Therefore: From every tensor a new one may be formed by differentiation with respect to x^0.

2. Covariant differentiation with respect to x^a: If ψ is a scalar then we have for cut-transformations:

$$\frac{\partial \psi}{\partial \bar{x}^a} = \frac{\partial \psi}{\partial x^a} - \frac{\partial \psi}{\partial x^0} f_{,a}. \tag{26a}$$

From this equation and (13), by eliminating $f_{,a}$, we get that

$$\frac{\partial \psi}{\partial x^a} - A_a \frac{\partial \psi}{\partial x^0}$$

is invariant with respect to the cut-transformations. Therefore the operator

$$\left(\frac{\partial}{\partial x^a} - A_a \frac{\partial}{\partial x^0} \right), \tag{26}$$

acting on a tensor, leaves its components invariant with respect to cut-transformations.

On the other hand, we know from the absolute differential calculus that the expression

$$B_{s,a} - \tfrac{1}{2}B_l g^{lm}(g_{ms,a} + g_{am,s} - g_{as,m}) \tag{27a}$$

(B_s being an arbitrary covariant vector) has a tensor character with respect to four- (but not cut-) transformations. By replacing $\frac{\partial}{\partial x^a}$ in this expression everywhere by the operator (26) we add new members which again have a tensor character with respect to four-transformations, this being true for $B_{s,0}$, $g_{mn,0}$, A_s. The expression obtained in this way certainly has a tensor character with respect to four-transformations. Since it is also invariant with respect to cut-transformations it satisfies our definition of a tensor.

We can, therefore, define an invariant differentiation in the following way:

$$B_{s;a} = B_{s,a} - A_a B_{s,0} - B_l \Gamma^l_{sa}, \tag{27}$$

where

$$\Gamma^l_{sa} = \tfrac{1}{2}g^{lm}[(g_{ms,a} - A_a g_{ms,0}) + (g_{ma,s} - A_s g_{ma,0}) - (g_{as,m} - A_m g_{as,0})]. \tag{27b}$$

In an analogous way we have for a contravariant vector with the same Γ^l_{sa}:

$$B^s_{;a} = B^s_{,a} - A_a B^s_{,0} + B^l \Gamma^s_{la}. \tag{28}$$

From this a corresponding rule for more general tensors follows immediately. The absolute differentiation defined thus satisfies the rule of differentiation of products.

By straightforward calculation one obtains

$$g_{mn;a} = g^{mn}{}_{;a} = 0. \tag{29}$$

3. The curvature tensor: By a straightforward calculation we find:

$$B_{s;a;b} - B_{s;b;a} = -B_l R^l_{.sab} - B_{s,0} A_{ab}, \tag{30}$$[6]

where

$$R^l_{.sab} = (\Gamma^l_{sa,b} - A_b \Gamma^l_{sa,0}) - (\Gamma^l_{sb,a} - A_a \Gamma^l_{sb,0}) - \Gamma^l_{am}\Gamma^m_{sb} + \Gamma^l_{bm}\Gamma^m_{sa}. \tag{30a}$$

From (30) follows, because of the tensor character of the last term, that $R^l_{.sab}$ is a tensor (the curvature tensor) which is antisymmetric with respect to its last two indices.

4. The commutation rule for the operators "$,0$" and "$;a$" is

$$(u_{m,0})_{;a} - (u_{m;a})_{,0} = u_s \Gamma^s_{am,0}. \tag{31}$$

The derivatives of the Γ^s_{am} with respect to x^0 have, therefore, tensor character

[6] $A_{ab} = A_{a,b} - A_{b,a}$.

Variational Principle and Field Equations

Starting from the formalism developed here we shall formulate the most general field equations satisfying the following conditions:

1. The field equations should be derived from a variational principle;

2. The action function should consist exclusively of terms containing either second derivatives linearly or products of two derivatives of the first order.

These are the tensors from which such an action function could be built algebraically:

$$R^{i}_{.klm}\ ,\ A_{mn}\ ,\ g_{mn}\ ,\ g_{mn,0}\ ,\ g_{mn,00}\ ,\ A_{mn,0}\ ,\ A_{mn;s}.$$

Some of the invariants formed from these tensors can, by partial integration of the Hamiltonian integral be reduced to the others. There remain only the following invariants which are, in this sense, independent of each other:

$$(32)\qquad \begin{cases} \underset{1}{H} = R^{i}_{.klm}\delta^{l}_{i}g^{km} = R_{km}g^{km} = R, \qquad \underset{2}{H} = A_{mn}A_{st}g^{ms}g^{nt} = A_{mn}A^{mn}, \\ \underset{3}{H} = g^{mn}{}_{,0}g_{mn,0}, \qquad \underset{4}{H} = g^{mn}g_{mn,0}g^{rs}g_{rs,0}. \end{cases}$$

The most general variational principle, satisfying our conditions formulated above, is therefore:

$$(33)\qquad \delta\int \mathfrak{H}dx^0dx^1dx^2dx^3dx^4 = 0, \qquad \mathfrak{H} = \sqrt{-g}\Big(\underset{1\,1}{\alpha H} + \underset{2\,2}{\alpha H} + \underset{3\,3}{\alpha H} + \underset{4\,4}{\alpha H}\Big),$$

where $\underset{1}{\alpha}\cdots\underset{4}{\alpha}$ are arbitrary constants. The variation has to be carried out with respect to the g^{mn} and the A_s. The integral should be taken over an arbitrary world region of the coördinates $x^1\cdots x^4$ and over one period of the coördinates x^0. The latter restriction is necessary because the δA_s must be chosen independent of x^0. (Compare (22)). Therefore the contributions from the boundary of the region do not enter into the usual operations performed on the action function only if the integral is taken over exactly one period of x^0. The calculations, which may be performed with the help of the methods quoted in the appendix, give:

$$(34)\qquad \begin{cases} \delta\int \underset{1}{\mathfrak{H}}dx^0\cdots dx^4 = \int \{[\tfrac{1}{2}R_{kl} + \tfrac{1}{2}R_{lk} - \tfrac{1}{2}g_{kl}R]\delta g^{kl} \\ \qquad\qquad + [g^{mn}\Gamma^{s}_{mn,0} - g^{ms}\Gamma^{n}_{mn,0}]\delta A_s\}\sqrt{-g}\,dx^0\cdots dx^4, \\ \delta\int \underset{2}{\mathfrak{H}}dx^0\cdots dx^4 = \int \{[2A_{km}A_{l.}^{\ m} - \tfrac{1}{2}g_{kl}A_{mn}A^{mn}]\delta g^{kl} \\ \qquad\qquad - [4A^{st}{}_{;t}]\delta A_s\}\sqrt{-g}\,dx^0\cdots dx^4 \\ \delta\int \underset{3}{\mathfrak{H}}dx^0\cdots dx^4 = \int [-2g_{kl,00} + 2g^{rs}g_{kr,0}g_{ls,0} - g^{rs}g_{rs,0}g_{kl,0} \\ \qquad\qquad - \tfrac{1}{2}g^{rs}{}_{,0}g_{rs,0}g_{kl}]\delta g^{kl}\sqrt{-g}\,dx^0\cdots dx^4, \\ \delta\int \underset{4}{\mathfrak{H}}dx^0\cdots dx^4 = \int g_{kl}[\tfrac{1}{2}(g^{mn}g_{mn,0})^2 + 2g^{mn}g_{mn,00} \\ \qquad\qquad + 2g^{mn}{}_{,0}g_{mn,0}]\delta g^{kl}\sqrt{-g}\,dx^0\cdots dx^4. \end{cases}$$

Here the δg^{kl} are arbitrary functions of the coördinates $x^0 \cdots x^4$. The sum of the coefficients of δg^{kl}, multiplied by their $\underset{i}{\alpha}$, must, therefore, vanish. However, the δA_s are, as we mentioned before, functions of $x^1 \cdots x^4$ only, but independent of x^0. Therefore, in this case, only the integral of the sum of the coefficients of δA_s, (multiplied by their $\underset{i}{\alpha}$), taken over one period of x^0, vanishes.

The fourteen field equations derived in this way from the variational principle (33), take the form:

$$(35)\quad \begin{cases} \underset{1}{\alpha}(\frac{1}{2}R_{kl} + \frac{1}{2}R_{lk} - \frac{1}{2}g_{kl}R) + \underset{2}{\alpha}(2A_{km}A_{l}^{\ m} - \frac{1}{2}g_{kl}A_{mn}A^{mn}) + \\ + \underset{3}{\alpha}(-2g_{kl,00} + 2g^{rs}g_{kr,0}g_{ls,0} - g^{rs}g_{rs,0}g_{kl,0} - \frac{1}{2}g^{rs}{}_{,0}g_{rs,0}g_{kl}) + \\ + \underset{4}{\alpha}g_{kl}[\frac{1}{2}(g^{mn}g_{mn,0})^2 + 2g^{mn}g_{mn,00} + 2g^{mn}{}_{,0}g_{mn,0}] = 0 = \dfrac{1}{\sqrt{-g}}\mathfrak{G}_{kl}, \end{cases}$$

$$(36)\quad \int \{\underset{1}{\alpha}(g^{mn}\Gamma^{s}_{mn,0} - g^{ms}\Gamma^{n}_{mn,0}) - 4\underset{2}{\alpha}A^{sl}_{;l}\}\sqrt{-g}\,dx^0 = 0 = \int \mathfrak{Z}^s\,dx^0.$$

Finally we shall derive the identities which the field equations satisfy. As it is customary, we shall find these identities by expressing analytically the invariance of the Hamiltonian integral with respect to an infinitesimal coördinate transformation. If we introduce for δg^{mn} and δA_s the variations produced by an infinitesimal coördinate transformation, then the variation of the integral vanishes. It should be noted that the variation is not to be carried out with the space points fixed, but with the coördinate values fixed. Denoting the infinitesimal coördinate transformations by

$$(37a)\quad \begin{cases} x^a = \bar{x}^a + \xi^a(\bar{x}^1 \cdots \bar{x}^4) \\ x^0 = \bar{x}^0 + \xi^0(\bar{x}^1 \cdots \bar{x}^4) \end{cases}$$

we obtain for δg^{mn} and δA_s:

$$(37b)\quad \begin{cases} \delta g^{kl} = -g^{sl}\xi^k_{,s} - g^{sk}\xi^l_{,s} + g^{kl}{}_{,s}\xi^s + g^{kl}{}_{,0}\xi^0, \\ \delta A_s = A_r\xi^r_{,s} + \xi^0_{,s} + A_{s,r}\xi^r. \end{cases}$$

These expressions can easily be brought into a form revealing more clearly their tensor character:

$$(37c)\quad \begin{cases} \delta g^{kl} = -g^{sl}\xi^k_{\ ;s} - g^{sk}\xi^l_{;s} + g^{kl}{}_{,0}(A_r\xi^r + \xi^0), \\ \delta A_s = (A_r\xi^r + \xi^0)_{;s} + A_{sr}\xi^r. \end{cases}$$

The expression $(A_r\xi^r + \xi^0)$ is an inner product of two five dimensional vectors, therefore a scalar. After performing some partial integrations (using the methods explained in the appendix) we obtain:

$$(37d)\quad \delta\int \mathfrak{H}dx^\alpha = \int \{[\mathfrak{G}_k{}^s{}_{;s} + A_{sk}\mathfrak{Z}^s]\xi^k + [\mathfrak{G}_{rs}g^{rs}{}_{,0} - \mathfrak{Z}^s_{;s}](A_r\xi^r + \xi^0)\}\,dx^\alpha = 0.$$

As ξ^k, ξ^0 are arbitrary functions of $x^1 \cdots x^4$, but independent of x^0 the integral (37d) will vanish identically if and only if the integrals of the square bracket expressions over one period of x^0 equal zero:

$$(37) \qquad \begin{cases} 2\int \mathfrak{G}_{k.}^{\ s}{}_{;s}\, dx^0 + A_{sk}\int \mathfrak{I}^s\, dx^0 = 0, \\ \int \mathfrak{G}_{rs} g^{rs}{}_{,0}\, dx^0 - \left\{\int \mathfrak{I}^s\, dx^0\right\}_{;s} = 0. \end{cases}$$

These are the identities of the field equations (35), (36).

The theory involves four universal constants, $\underset{2\ 1}{\alpha/\alpha}$, $\underset{3\ 1}{\alpha/\alpha}$, $\underset{4\ 1}{\alpha/\alpha}$, λ. The constant $\underset{2\ 1}{\alpha/\alpha}$ corresponds to the gravitational constant involving a connection between the (arbitrary) units of length and mass. One of the other constants, for instance λ, will depend on the unit of length. The remaining two are "genuine" universal constants which cannot be eliminated from the theory.

Summary

Kaluza's five dimensional theory of the physical space provides a unitary representation of gravitation and electromagnetism. It demands, if consistently interpreted as a five dimensional theory, the existence of a vector A, whose symmetrical covariant derivatives vanish in the whole space. Apart from its physical meaning, this assumption seems to be artificial. Furthermore the representation of a purely four dimensional continuum by a five dimensional one appears not to be sufficiently justified.

By our generalization we try to eliminate these objections. The hypothesis of cylindricity, (i.e. of the existence of a vector A with vanishing symmetrical derivatives) is replaced by the assumption that the space is closed in the direction of the fifth coördinate. By this change the basic assumptions of the theory are considerably simplified. Furthermore it is much more satisfactory to introduce the fifth dimension not only formally, but to assign to it some physical meaning. Nevertheless there is no contradiction with the empirical four dimensional character of physical space.

Appendix

Analysis of Tensor Densities

The tensor density plays, beside the concept of tensors, an important rôle in relativistic investigations. Although the concept of tensor densities can easily be reduced to that of tensors nevertheless the whole formalism and especially the derivation of the field equations from the variational principle are considerably simplified if tensor densities are introduced independently. Although the analytical properties of tensor densities have been fully investi-

gated[7] we shall, in order to make the reading easier, make some remarks on the subject.

A tensor density differs from a tensor by a factor (common to all components) having the transformation character of a scalar density. A scalar density $\underset{(n)}{\rho}$ of the "weight" n is a quantity with the following transformation property:

$$\underset{(n)}{\bar{\rho}} = \underset{(n)}{\rho} \left| \frac{\partial x^\alpha}{\partial \bar{x}^\beta} \right|^n . \tag{A 1}$$

There exist, correspondingly, tensor densities of different weights. (Ordinary tensors are tensor densities of the weight zero.) By multiplying two tensor densities of the weights m and n we obtain a tensor density of the weight $(m + n)$. The rules of a tensor density algebra follow immediately from that of the tensor algebra and need not be derived here.

For the sake of simplicity we shall consider here only quantities in a four dimensional continuum; the results are, however, independent of the number of dimensions.

There exist, in every four dimensional space, the two tensor densities of Levi-Civita

$$\underset{(-1)}{\delta_{iklm}}$$

and

$$\underset{(+1)}{\delta_{iklm}},$$

antisymmetric in all indices, the components of which are $+1$ or -1, depending of whether $(iklm)$ is an odd or an even permutation of (1234). Their tensor density character follows immediately from the transformation law of tensor densities. From the Levi Civita tensor densities the well known Kronecker tensor $\delta_m^{m'}$ (with components equal 1 if $m = m'$ and 0 if $m \neq m'$) can be obtained by:

$$\delta_m^{m'} = \tfrac{1}{6} \underset{(-1)}{\delta_{iklm}} \underset{(+1)}{\delta^{iklm'}} . \tag{A 2}$$

We assume in the space under consideration the existence of symmetrical components of parallel displacement (the space functions $\Gamma^l_{ik} = \Gamma^l_{ki}$) by which the covariant derivatives of a tensor are defined in the usual way:

$$B^{k\cdots}_{i\cdots;s} = B^{k\cdots}_{i\cdots,s} + B^{l\cdots}_{i\cdots}\Gamma^k_{ls} + \cdots - B^{k\cdots}_{l\cdots}\Gamma^l_{is} - \cdots . \tag{A 3}$$

This definition of the covariant differentiation has two essential properties:

1. It satisfies the rule for the differentiation of products of the ordinary calculus:

$$(AB)_{;s} = AB_{;s} + BA_{;s} .$$

[7] V. Hlavatý, Annalii di Mathematica, V, Ser. IV (1927/28): *Théorie des densités dans le displacement général.*

2. The absolute derivatives of the Kronecker tensor vanish identically:

$$\delta^k_{i;s} = 0\,.$$

We shall now define the covariant derivative of a tensor density so that:

1. The rule for the differentiation of products is preserved;

2. The absolute derivatives of the co- and contravariant tensor densities of Levi Civita vanish identically.

By direct calculation we confirm that these demands are satisfied by the following definition:

$$\underset{(n)}{B^{k\cdots}_{i\cdots;s}} = \underset{(n)}{B^{k\cdots}_{i\cdots,s}} + \underset{(n)}{B^{l\cdots}_{i\cdots}}\Gamma^k_{ls} + \cdots - \underset{(n)}{B^{k\cdots}_{l\cdots}}\Gamma^l_{is} - \cdots - n\underset{(n)}{B^{k\cdots}_{i\cdots}}\Gamma^r_{sr}. \tag{A 4}$$

This is the rule for the differentiation of tensor densities of the weight n.

If we specialize the space further by assuming that it is Riemannian then the components of the parallel displacement are given by:

$$\Gamma^s_{mn} = \tfrac{1}{2}g^{st}(g_{ml,n} + g_{nl,m} - g_{mn,l}), \tag{A 5}$$

and therefore

$$\Gamma^n_{sn} = \frac{1}{2}\frac{g_{,s}}{g}, \qquad g = |g_{ik}|. \tag{A 6}$$

The determinant g is a scalar density of the weight 2. By applying (A 4) we get:

$$\underset{(2)}{g_{;s}} = g_{,s} - 2g\left(\frac{1}{2}\frac{g_{,s}}{g}\right) = 0. \tag{A 7}$$

Therefore the absolute derivative of $\sqrt{g}$ and, because of the rule for the differentiation of products, the absolute derivatives of $g_{mn}\sqrt{g}$ and of $g^{mn}\sqrt{g}$ vanish also.

We shall now show how the tensor densities can be applied in the variational process.

1. The derivation of the equations of gravitation by the variation of Riemann's curvature tensor:

$$\begin{cases} \delta\int \underset{(1)}{\mathfrak{H}}\, d\tau = 0, \\ \underset{(1)}{\mathfrak{H}} = \sqrt{-g}\,g^{km}\delta^l_i R^i_{.klm}. \end{cases} \tag{A 8}$$

It was shown by Palatini that the variation with respect to the g^{km} can be simplified, because of the equation:

$$\delta R^i_{.klm} = (\delta\Gamma^i_{kl})_{;m} - (\delta\Gamma^i_{km})_{;l} \tag{A 9}$$

which can easily be verified (note that $\delta\Gamma^i_{kl}$ is a tensor!). Because of

$$\delta(\sqrt{-g}\,g^{km}) = \sqrt{-g}(\delta g^{km} - \tfrac{1}{2}g^{km}g_{ab}\cdot\delta g^{ab}), \tag{A 10a}$$

we have then

$$\delta \int \underset{(1)}{\mathfrak{H}} d\tau = \int \sqrt{-g}(R_{km} - \tfrac{1}{2} g_{km} R)\delta g^{km} d\tau + \int \sqrt{-g} g^{km}[(\delta\Gamma^l_{kl})_{;m} - (\delta\Gamma^l_{km})_{;l}] d\tau. \tag{A 10b}$$

As $(g^{km} \cdot \sqrt{-g})_{;l}$ vanishes the integrand of the second integral may be written

$$[\sqrt{-g}(g^{kl}\delta\Gamma^m_{km} - g^{km}\delta\Gamma^l_{km})]_{;l}, \tag{A 10c}$$

where the square bracket expression $\underset{(1)}{K^l}$ is a vector density of the weight 1. It follows, therefore, from (A 4) that $\underset{(1)}{K^l_{;l}}$ may be replaced by $\underset{(1)}{K^l_{,l}}$. This integral vanishes, according to Gauss' theorem, if the variations of the g_{ik} and of the Γ^l_{ik} vanish on the boundary of the integration region. There remains only the first integral of (A 10b) so that the result is

$$R_{km} - \tfrac{1}{2} g_{km} R = 0. \tag{A10}$$

2. Tensor densities in the space of the generalized Kaluza theory. In such a space, and in the special coördinate system, a scalar density of the weight n can be defined by the transformation law:

$$\underset{(n)}{\bar{\rho}} = \underset{(n)}{\rho} \left| \frac{\partial x^\alpha}{\partial \bar{x}^\beta} \right|^n \quad \text{for four-transformations,}$$
$$\underset{(n)}{\bar{\rho}} = \underset{(n)}{\rho} \quad \text{for cut-transformations.} \tag{A 11}$$

The transformation law of tensor densities is also determined by this definition, i.e. reduced to that of tensors.

Covariant differentiation of tensor densities: We can expect, because of (A 4) and (27), (28), that the following formula expresses the differentiation rule:

$$\underset{(n)}{B^{k\cdots}_{i\cdots;s}} = \underset{(n)}{B^{k\cdots}_{i\cdots,s}} - A_s \underset{(n)}{B^{k\cdots}_{i\cdots,0}} + \underset{(n)}{B^{l\cdots}_{i\cdots}}\Gamma^k_{ls} + \cdots - \underset{(n)}{B^{k\cdots}_{l\cdots}}\Gamma^l_{is} - \cdots - n\underset{(n)}{B^{k\cdots}_{i\cdots}}\Gamma^l_{sl}. \tag{A 12}$$

$\underset{(-1)}{\delta_{iklm}}$ is a tensor density with respect to four-transformations and invariant with respect to cut-transformations. It is, therefore, a tensor density according to our definitions. The absolute derivative of this tensor density, given by (A 12), vanishes. It can also be seen that (A 12) preserves the rule for the differentation of products.

It follows from our definitions that $\underset{(2)}{g} = |\, g_{mn}\,|$ is a scalar density of the weight 2 with the absolute derivatives:

$$\underset{(2)}{g_{;s}} = g_{,s} - A_s g_{,0} - 2g\Gamma^l_{sl} = 0. \tag{A 13}$$

(This vanishes because of (27b).) Furthermore the equation holds:

(A 14) $$g_{mn;s} = 0.$$

We form the divergence $\underset{(1)}{B^s_{;s}}$ of a vector density $\underset{(1)}{B^s}$ of the weight 1. According to (A 12) we have

(A 15a) $$\underset{(1)}{B^s_{;s}} = \underset{(1)}{B^s_{,s}} - A_s \underset{(1)}{B^s}_{,0},$$

and as A_s does not depend on x^0

(A 15) $$\underset{(1)}{B^s_{;s}} = \underset{(1)}{B^s_{,s}} - (A_s \underset{(1)}{B^s})_{,0}.$$

It is essential to note here that both terms on the right hand side take the form of ordinary derivatives.

We have now, as in the gravitational theory, to form the variation of the integral (Compare (32), (33))

(A 16a) $$\int \sqrt{-g}\, g^{km} \delta^l_i R^i_{.klm}\, d\tau.$$

This integral, however, is to be taken over a region of the variables $x^0, x^1 \cdots x^4$, which contains exactly one period of the coördinate x^0 (and is arbitrary with respect to $x^1 \cdots x^4$). The $R^i_{.klm}$ are defined by (30a). The variation of this quantity takes the form:

(A 16b) $$\delta R^i_{.klm} = (\delta\Gamma^i_{kl})_{,m} - (\delta\Gamma^i_{km})_{;l} - \delta A_m \Gamma^i_{kl,0} + \delta A_l \Gamma^i_{km,0}.$$

The variation of this integral gives, as far as the variation of the factor $g^{km} \cdot \sqrt{-g}$ is concerned, the contribution:

(A 16c) $$\int \sqrt{-g}(\tfrac{1}{2}R_{km} + \tfrac{1}{2}R_{mk} - \tfrac{1}{2}g_{km}R)\delta g^{km}\, d\tau.$$

The variation of $R^i_{.klm}$ gives, because of the two last terms in (A 16b),

(A 16d) $$\int \sqrt{-g}(g^{km}\Gamma^l_{km,0} - g^{kl}\Gamma^m_{km,0})\delta A_l\, d\tau.$$

The first two terms on the right hand side of (A 16b), however, do not give any contribution. The proof is similar for both terms; we shall, therefore, show it only for the first one. The integrand

$$\sqrt{-g} \cdot g^{km}(\delta\Gamma^l_{kl})_{;m}$$

can, because of

$$(\sqrt{-g} \cdot g^{km})_{;m} = 0,$$

(following from (A 13), (A 14)) be brought into the form:

$$\underset{(1)}{B^m_{;m}}.$$

But such an integral vanishes because of (A 15). It does so partially because the variations vanish on the x^a-boundaries, partially because of the periodicity of B^m with respect to x^0. The final result of the variation is:

$$\text{(A 16)} \quad \int \sqrt{-g}\{(\tfrac{1}{2}R_{km} + \tfrac{1}{2}R_{mk} - \tfrac{1}{2}g_{km}R)\delta g^{km} + (g^{km}\Gamma^{l}_{km,0} - g^{kl}\Gamma^{m}_{km,0})\delta A_l\}\, d\tau,$$

where the δg^{km} depend arbitrarily on x^a, x^0, the δA_s however only on x^a.

The calculations applied in the derivation of the identities are quite similar.

INSTITUTE FOR ADVANCED STUDY.